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 T_{soil} and the difference of T_{soil}-T_{air} 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 km^2, which is comparable with the IPA map estimate (16.2 million km^2) (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 km^2) and higher (20.01 million km^2) 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 km^2). 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 km^2). 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 T_{air}-d_{snow}-T_{soil} 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 T_{soil} and T_{soil}- T_{air} to d_{snow} 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 T_{soil} bias table:

T_{soil} bias statistics (n=479); StDev^{Obs}=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 T_{soil} in comparison to station observations. For example, both figures clearly show the strongest underestimation of T_{soil} 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 T_{soil} is biased, the relationship between T_{soil} and T_{air} can be well represented compared with observations. For example, CLM45 and JULES have a cold bias in T_{soil} (too cold by ca. 3 K), but can represent the dependency of the T_{soil}-T_{air} 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 T_{soil} and the difference of T_{soil}-T_{air} 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 ΔT 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 ΔT 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 T_{soil} and T_{air}. Gradient between T_{soil} and T_{air} 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 T_{soil} and

T_{air} 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 T_{air} vs.T_{soil} 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 T_{air} vs.T_{soi} 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 ℃. These models better reproduce the observed ΔT vs. d_{snow} 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-T_{soil} relation to snow conditions (Fig.4, Tab.2). Another measure quantitatively confirms the same models behavior: The observed average d_{snow} 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 T_{soil} vs T_{air} 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 d_{snow} relationships (ISBA, UVic) also fail in this evaluation of their T_{soil} vs T_{air} relationship. They simulate a too strong sensitivity of T_{soil} to T_{air} (regression slopes larger than 0.9 ℃/ ℃, R^2>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 T_{soil} and T_{air} 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 T_{soil} vs T_{air} relationships from observations reaches ca. 10 ℃.

2) We agree and we changed the wording through all the manuscript. We use either "larger slope in the regression between T_{soil} and T_{air}" or "larger regression slope", or "stronger sensitivity of T_{soil} to T_{air}".

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)*."

<u>Minor points</u>

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 "(ΔT)".
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 T_{soil} on T_{air} for three T_{air} ranges".
Done. It reads now "*We illustrate the dependence of ΔT and T_{soil} on d_{snow} for three T_{air} 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 ΔT and d_{snow} 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: $ΔT/d_{snow}$ do mean a ratio as shown in Table 2. Revise "$ΔT/d_{snow}$ relationship", to "$ΔT$-d_{snow} 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 ΔT/dsnow relationship in the complementary form of the PDFS of …", revised as "Figure 2 shows the ΔT-dsnow relationship in a complementary form using the PDFS of ……"
Done. We also change "ΔT-d_{snow}" to "ΔT vs. d_{snow}". 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 T_{soil} under thick snow conditions”. Probably should be “emphasizing the reduced sensitivity of T_{soil} to snow depth under thick snow conditions”
Done.

Figures: revise “AirT” to “T_{air}”
Done.

<u>Response to comments of reviewer #2:</u>

We thank Reviewer 2 for providing valuable comments on our paper 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*.

<u>Comments:</u>

1. P5 L22-24: Most of the analysis is based on using the meteostation data across the Russian territory from 1980-1990. Unfortunately, there is no in-depth analysis of the observations and why these data can be used. Observations at many meteostations are performed at the disturbed conditions and thus only a small subset of this data could be used in the direct comparison. The authors simply assume that just all data could be used. The meaningful comparison of the model to observations at the disturbed sites could be done if the model also simulates the disturbed conditions. One of the major problems in this paper is that many observations were taken at the disturbed sites near meteostations, while models simulate typical ground conditions. To do the comparison correctly it is either necessary to model disturbed conditions in the models or remove 'corrupted' observations from the analysis.

This data set is quality checked and officially released by the All-Russian Research Institute of Hydrometeorological Information-World Data Centre (RIHMI-WDC; http://meteo.ru/). They provide quality controlled soil temperature at depths to 320 cm from meteorological stations of the Russian Federation. The data was subject to quality control by using four methods of statistical control (histograms, standard deviation check, check for connectedness of values that are neighboring in time).

Meteorological sites are located in an open and typical place within the surrounding terrain. They are located far from the major obstacles and water bodies that may have a direct effect on the quality of measurements. To keep the surface of the meteorological site in its natural state, it is permitted to walk only on the specially laid tracks, a width not exceeding 40 cm, within the site territory. In the warm season grass on the site is regularly mowed and trimmed. The height of the grass on the site is not more than 20 cm. In the winter, the natural state of the snow cover is not broken. Observing conditions at the Russian stations in all meteorological elements correspond with WMO standards. The observations presented have been included in data sets, such as GSOD, HadSRUT4 etc. and are widely used in climate research.

Soil temperature measurements are carried out simultaneously with the measurements of the whole complex of meteorological observations (temperature, characteristics and dynamics of snow cover and precipitation and so on). The soil temperature observations are under the original surface (from draw-out thermometer data). All meteorological observations, including soil temperature, are produced exclusively within the same site area (26 by 26 meters in size), and under constant careful control of workers who are caring for a site. This prevents any accidental violations of the integrity of the site and guarantees the quality of observations. A detailed history of the development of methodology for soil temperature measurements was provided by Bykhovets et al. (2007). A detailed description of dataset preparation is provided in Sherstiukov (2007). The archived soil temperature dataset was run through four independent methods of quality control (Sherstiukov, 2012).

Thus, the data of meteorological observations at the Russian stations for soil temperature can be effectively used in the evaluation of thermal changes in the upper layers of permafrost zone, as well as in the analysis of processes of thermal interaction between the atmosphere and soil. Similar conclusions were obtained by leading Russian scientists in this area (Anisimov and Sherstiukov, 2016; Pavlov and Malkova, 2009) and internationally (Park et al., 2014; Brun et al., 2013; Decharme et al. 2016; PaiMazumder et al., 2008).

However, as Park et al. (2014) pointed out, the observations at some locations could have been unavoidably disturbed by grass cutting during the warm season and the removal of organic materials, mainly at agricultural sites. These disturbances may cause increased warming of the soil over time. Therefore, long-term soil temperature trends in could potentially include this non-climatic component (Frauenfeld et al 2004).

Thus, we agree that we have to be careful which and how we use the observation data. But we can argue that our study results are solid. First, and most important is that we do not present either direct point-by-point comparison with station observations nor trends, but we investigate variables relationships. We evaluate the models functional behaviors (inter-variable scatter plots, Figs. 1, 4, 5 and PDFs, Fig. 2). The inter-variable scatter plots (which show medians and the 25th and 75th percentiles) as well as the PDFs present the overall functional behavior of snow insulation effects, and not the individual station's representation. Therefore, this kind of benchmark to evaluate the models skill is a solid approach using the station data, which include uncertainties (by showing the percentiles). This approach, as the other reviewer points out, reveals some structural issues of the models in simulating snow depths and its insulation effects on soil temperature. Secondly as we are only looking at upper soil temperatures, any change due to changes in soil column temperature gradient resulting from grass removal are negligible. Thirdly, the agriculture sites concerns are not an issue for our permafrost area focused study.

According to the reviewer's comment, we include in section 2.2 two references for more details of the data set: "*A detailed description of dataset preparation is provided in Sherstiukov (2012a). Observing conditions at the Russian stations in all meteorological elements correspond with WMO standards. The observations presented have been included in data sets, such as GSOD, HadSRUT4 etc. and are widely used in climate research (e.g. Anisimov and Sherstiukov, 2016; Decharme et al. 2016; Park et al., 2014; Brun et al., 2013; Pavlov and Malkova, 2009; PaiMazumder et al., 2008).The soil temperature dataset was run through four independent methods of quality control (Sherstiukov, 2012b).*"

Further, we emphazise the point that possible disturbances do not affect our results:

"*However, some soil temperature observations could be disturbed by grass cutting during the warm season and the removal of organic materials, mainly at agricultural* *site, which may affect the trend in warm season (Park et al., 2014), but this does not affect our results about the air- upper soil temperature relationship in winter*."

Finally, we cannot compare the model results and observation over only bare ground (or short grass) in the PCN simulations. Only the grid-cell average results were kept (and not the value for each vegetation and bare ground tile).

References:

Anisimov, O.A., Sherstiukov A.B. Evaluating the effect of environmental factors on permafrost factors in Russia, Earth's Cryosphere, XX(2), 90-99, 2016.

Brun, E., Vionnet, V., Boone, A., Decharme, B., Peings, Y., Valette, R., Karbou, F. and Morin, S.: Simulation of northern Eurasian local snow depth, mass and density using a detailed snowpack model and meteorological reanalysis, J. Hydrometeorol., 14, 203–214, doi:10.1175/jhm-d-12-012.1, 2013.

Bykhovets, S. S., Sorokovikov, V. A., Martuganov, R. A., Mamykin, V. G., and Gilichinsky, D. A. History of soil temperature measurements at the network of meteorological stations in Russia, Earth's Cryosphere, XI(1), 7-20, 2007.

Decharme, B., Brun, E., Boone, A., Delire, C., Le Moigne, P. and Morin, S.: Impacts of snow and organic soils parameterization on North-Eurasian soil temperature profiles simulated by the ISBA land surface model, The Cryosphere, 10, 853–877, doi: 10.5194/tc-10-853-2016, 2016.

Frauenfeld O W, Zhang T, Barry R G and Gilichinsky D.: Interdecadal changes in seasonal freeze and thaw depths in Russi,a J. Geophys. Res. 109 D05101, 2004

Park, H., Sherstiukov, A.B., Fedorov, A.N., Polyakov, I. V., Walsh, J.E.: An observation-based assessment of the influences of air temperature and snow depth on soil temperature in Russia, Environmental Research Letters, Vol. 9, 2014, http://iopscience.iop.org/1748-9326/9/6/064026

PaiMazumder, D., Miller, J., Li, Z., Walsh, J. E., Etringer, A., McCreight, J., Zhang, T., Mölders, N. Evaluation of Community Climate System Model soil temperatures using observations from Russia, Theoretical and Applied Climatology , 94(3),187-213, 2008.

Pavlov, A.V., Malkova, G.V. Small-scale mapping of trends of the contemporary ground temperature changes in the Russian North, Earth's Cryosphere, XIII(4), 32-39, 2009.

Sherstiukov, A. Dataset of daily soil temperature up to 320 cm depth based on meteorological stations of Russian Federation, RIHMI-WDC, 176, 224-232, 2012a.

Sherstiukov, A. Statistical quality control of soil temperature dataset, RIHMI-WDC, 176, 224-232, 2012b.

2. P5 L29: "comprehensive Russian station data set". Please describe this comprehensive dataset. Provide a reference and in-depth discussion about the site conditions, disturbances of the ground cover, what stations are not qualified for the comparison.

This comment is related to the comment above. Please see our answer above.

3. P7 L26: "Snow depth was then calculated from SWE using a snow density of

250 kg m^{-3}." What was the rational to use 250 kg/m^3. Please provide references. How would the results change is 300 kg/m^3 is used? Why not to use the SWE instead of the snow depth in all other further comparisons. It looks like converting to the snow height can add additional uncertainties to the consecutive analysis.

The stations only give the snow depth (and not SWE). In addition to the station's ground snow observations we also use the gridded SWE data from the GlobSnow-2 product to support the station data results. Thus, to compare the GlobSnow data with the station data, we must convert one data set. No way has a preference. We decided to convert SWE from GlobSnow to snow depth, and for this we need an assumption about snow density. We use 250 kg m^{-3} because it is a mean observed value. Zhong et al. (2013) report snow density values of 180-250 kg m^{-3} for tundra/taiga and 156-193 kg m^{-3} for alpine snow classes in winter. Woo et al. (1983) report snow density values of 250-400 kg m^{-3} for various terrain types. Thus we use the mid-value of 250 kg m^{-3}. We added a paragraph in section 2.2: "*Snow depth was then calculated from SWE using a snow density of 250 kg m^{-3}, which is a median observed value in winter. Zhong et al. (2013) report snow density values of 180-250 kg m-3 for tundra/taiga and 156-193 kg m-3 for alpine snow classes. Woo et al. (1983) report snow density values of 250-400 kg m-3 for various terrain types. Choice of density does not materially affect the results.*"

To show you that this parameter choice does not affect our results, Fig. X1 is the comparison with snow depth from GlobSnow using a density of 300 kg m^{-3}. This figure confirms that the pattern does not change and shows that the differences in snow depth are small (less than 10 cm).

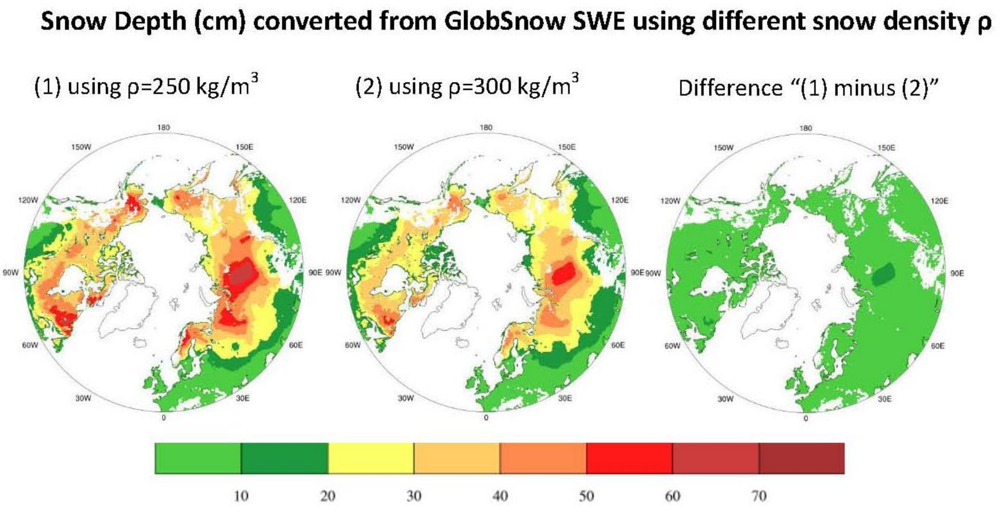


Figure X1: Snow depth (cm) derived from GlobSnow SWE using different assumptions of snow density, winter 1980-2000.

Further, our calculations show that the model biases in snow depth are quite similar using either the station or the GlobSnow data (see earlier SI Tab.3, now moved to Table 3). Also, if we re-calculate the model biases with respect to GlobSnow derived with a density of 300 kg m-3, the biases change only slightly and the model ranking is unaltered. For your convenience, we list here the comparison of the RMSE for snow depth (cm):

RMSE of snow depth (cm)			
	GlobSnow		Stations
	Density=250 kg/m^3	Density=300 kg/m^3	
CLM4.5	18.1	20.5	18.1
CoLM	22.1	24.9	21.4
ISBA	19.8	22.9	18.8
JULES	12.8	11.5	14.1
LPJ-GUESS	16.0	13.7	17.3
MIROC-ESM	14.0	13.7	17.9
ORCHIDEE	15.3	12.4	16.5
UVic	16.8	16.5	18.9
UW-VIC	20.0	22. 6	19.8

4. P7 L29: "the GlobSnow product can show regional differences (of ca. 0.5-5 cm) with biases increasing with increasing SWE" Is it a systematic bias in conversion of the SWE to snow height? If so, why not to take it into the account. Please specify where the regional differences are the largest, what areas have smallest. How many stations are located in the areas where differences are largest. Also, please list differences in % not the absolute values, since in some locations 5cm could be 25% of error vs 10% of error.

We do not aim to evaluate the GlobSnow data. There are quite a lot of papers about details of GlobSnow validation and uncertainties. The given sentences are from the references. The paper is already rather long and we do not feel a digression into this data set is of interest to most readers. Again we emphasize that all the relationship analysis we do is based on the station data (Figs. 1-5). We use GlobSnow only for the evaluation of the simulated spatial maps of snow depth (Fig.6). This has been done to support the station data results, and to arrive at a more solid comparison with the models. And indeed, we show that the calculated model biases in snow depth are quite similar using either the station or the GlobSnow data (see earlier SI Tab.3; now *Table 3*). To make this clearer in the text, we now directly cite the sentence from the references and also include a reference related to our considered Russian Arctic or Eurasia region. We improved the according part in section 2.2: *"…GlobSnow-2 product (http://www.globsnow.info/swe/), which has been produced using a combination of passive microwave radiometer and ground-based weather station data (Takala et al., 2011). Orographic complexity, vegetation cover, and snow state (e.g. wet snow) affect the accuracy of this product. When compared with ground measurements in Eurasia, the GlobSnow product can show root-mean-square error (RMSE) values of 30 to 40 mm for SWE values below 150 mm, with retrieval uncertainty increases when SWE is above this threshold (e.g., Takala et al., 2011; Muskett, 2012; Khelemet et al., 2013)."*

5. P8 L11-13: "We assume that there is relatively little impact due to soil moisture and texture between surface and 20 cm depth in winter." This is probably true for the cold climate conditions, while in the warm climate the surface might freeze, but 0.20m could stay thawed.

Yes, we consider here cold climate winter conditions. We follow the other reviewer's suggestion to formulate it "*In winter, the effects of other factors (e.g. soil moisture, texture) on ΔT are much smaller than that of snow*."

6. P8 L13-14: "Although we recognize the difference between ground surface and 20 cm soil temperatures and that soil organic layer could play a role in certain locations". What certain locations? In warm climate? Please be specific.

We do agree that the snow effect is the key in winter cold climate. You (and the other reviewer) also emphasize, other factors like soil moisture, texture, organic matter are of secondary importance, and we agree on this. Therefore, we deleted this sentence to avoid confusion.

7. P8 L17: "However, we find that the results do not significantly change when the model simulated temperature differences between ground surface temperature (GST) and near-surface air temperature (Tair) are used instead of between 20 cm soil (T20cm) and near-surface air temperatures." Again please be specific. How much is not significant?

This sentence is to confirm that our results of the functional behavior (Figs. 1-4) do not qualitatively change if we use either the ΔT=Tsoil-Tair or ΔT=GST-Tair. It is to tell the reader that we checked this. To make it clear, we do not have GST observations! Therefore, we have to use Tsoil. The question is how sensitive the results would be when using GST instead of upper soil-layer temperature (Tsoil at 20 cm depth). This would give an indication if/how soil characteristics are important for our presented relationships. The only thing we can do for this is to look into the model results. For your convenience, we show here one example plot (Fig. X2). The comparison of the red (ΔT=Tsoil-Tair) and green (ΔT=GST-Tair) clearly show that the regression does not materially change. "*To test how sensitive are results using 20 cm temperatures instead of ground surface, we also analysed model simulated temperature differences between ground surface and Tair, and found not qualitative differences, hence justifying use of 20 cm observations.*"

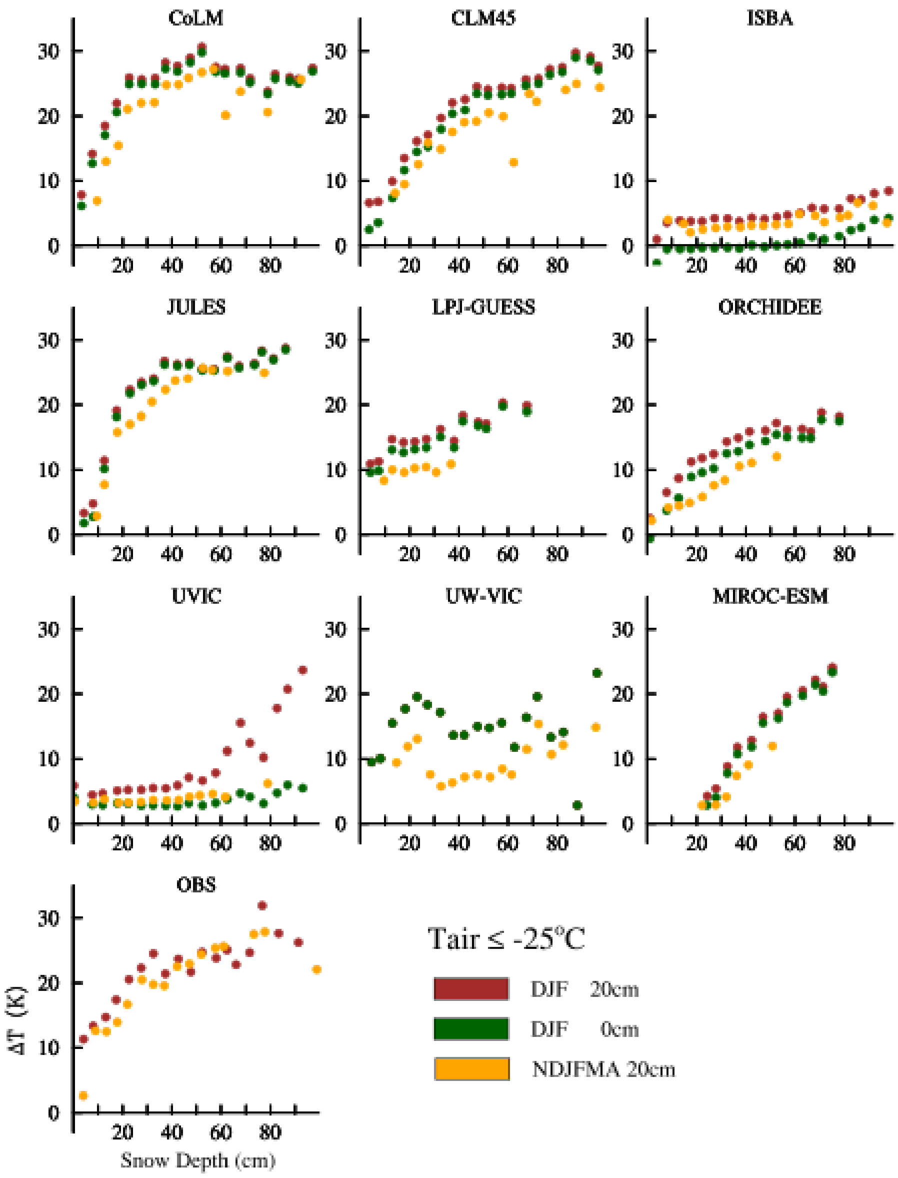


Figure X2: Variation of ΔT (K) with snow depth (cm) under cold conditions (Tair ≤ -25 ℃) for winter 1980-2000. The dots represent the medians of 5 cm snow depth bins, calculated from all Russian station grid points (n=268) and 21 individual winters. Red: ΔT is the difference between soil temperature at 20 cm depth and Tair, winter (DJF); Green: ΔT is the difference between ground surface temperature and Tair, winter (DJF); Orange: ΔT is the difference between soil temperature at 20 cm depth and Tair, winter (NDJFMA).

8. P8 L4: "we checked that a different winter definition (NDJFMA) does not substantially change the results" What is substantial, please be specific.

This sentence is to confirm that our results of the functional behavior (Figs. 1-4) do not change if we use another definition for the winter season. It is to tell the reader that we checked this. The question is how sensitive the results would be when using a different winter definition, because we know that snow can begin in November and end at the beginning of May. For your convenience, we show here one example plot

(Fig. X2). The comparison of the red (winter-DJF) and orange (winter-NDJFM) clearly show that the regression does not materially change. We deleted "substantially" and substitute "qualitatively". "*Our analysis is focused on the common winter (DJF) condition, although snow can begin in November and end at the beginning of May, but we checked that a different winter definition (NDJFMA) does not qualitative change any of the inter-variables relationships found*."

P3 L5-6: delete "as expressed by simulated differences"
Done.

P4 L5: Please try to avoid parentheses, usually it is possible to write the manuscript without them.
Done.

P4 L7-8: delete "quality" and parentheses.
Done.

P4 L14, "soil temperature", reference is missing.
The reference is provided just in the former sentence.

P4 L22-23: Please rewrite without parentheses.
Done.

P8 L21: "We use correlation analysis to investigate the co-variability…". Please provide formulae or references to this analysis
Done. "*We use the Pearson product-moment correlation coefficient and its significance (von Storch and Zwiers, 1999)…* "

P8 L22: "The input consists of detrended time series of winter means at each grid point." How did you compute detrended time series? You may add details to the appendix.
Done. "*Before we compute the correlations we detrended the data by removing a least squares regression line*."

P8 L26: "Student t-test" Reference?
Done. "*… coefficients is estimated by the Student's t-test (von Storch and Zwiers, 1999).*"

P8 L28: "To further examine the functional behavior between different variables, we present relationship…". What variables? Please be specific.
Done. We introduce the key variables in the paragraph above. "*… 4 key variables: near-surface air temperature (T_{air}), near-surface soil temperature (soil temperature at 20 cm depth; T_{soil}), snow depth (d_{snow}), and the difference between T_{soil} and T_{air}.*"

P8 L31: "We illustrate the dependence on air temperatures by evaluating". Dependence of what? What variable? Please be specific

Done. "*We illustrate the dependence of ΔT vs. dsnow and T_{soil} vs. d_{snow} relations for three Tair ranges*."

P9 L2-3: "The principal motivation for such classifications is to distinguish dry snow pack regimes from those where sporadic melt may occur even in winter." Usually, the motivation or idea to do something goes first. The next sentence describes how this motivation is implemented. Please re-write this paragraph such that the motivation is at the beginning.

Done. We revised this according to your suggestion. We start with "*To distinguish dry snow pack regimes from those where sporadic melt may occur even in winter, we split Tair into 3 regimes: the coldest conditions…* "

P9 L11: "observations and all models produce a clear relationship" I do not see this "clear relationship" at the first glance in Figure 1. Please be specific. Describe what the reader may find on this figure and then state the relationship.

Done. Fig. 1 clearly indicates that the observations as well as all models show an increase of ΔT with increasing dsnow. We improved this paragraph. First we start with the observations "*The air-soil temperature difference (ΔT) - snow depth (d_{snow}) relationship in winter (Fig. 1) shows in the Russian station observations an increase of ΔT with increasing d_{snow}. The data exhibit a linear relation between ΔT and d_{snow} at relatively shallow snow depths with a trend towards asymptotic behavior at thicker snow, which is in agreement with earlier findings (Zhang, 2005; Ge and Gong, 2010; Morse et al., 2011)…*". Then we continue with the models behavior "*All models reproduce the observed relationship, i.e. increasing ΔT with increasing d_{snow}. However, Fig. 1 also shows a wide across-model spread in the simulated relationships, and that some of the models are not consistent with the behavior in the observations…*"

P9 L14: "that some of the models are not consistent with the behavior in the observations. There is also significant scatter in the observation-based relationship, the inter-quartile range…". Observations are just plotted on the last panel in figure 1. It takes time to find the observations on figure 1. Please try to re-design Figure 1 and all other figures such that observations stand out and could be easily noted. And please point out/circle this scatter on the figure.

Sorry, we do not agree that it is hard to recognize the observations in the figures, though obviously this a personal issue. However, we consistently show observations in the bottom subpanel of our figures. To highlight the scatter in addition to the median (which is presented by the dots) we plot therefore upper and lower bars on these dots, indicating the 25th and 75th percentiles.

P9 L21-24: "The Russian station data and some model results exhibit a linear relation...larger snow depths (Fig. 1), which is in agreement with earlier findings (Zhang, 2005; Ge and Gong, 2010; Morse et al., 2011)." Move this sentence up to

the beginning of the above paragraph.
Done. We largely revised the discussion in Section 3.1, such that we discuss observations first, and then followed by the model representation.

P9 L31-32: "Figure 2 views the…relationship in the complementary form of the PDFs …different snow depth and air temperature regimes". Before showing the PDF, please explain what you plan to analyze and how the PDF can help you to achieve this goal. It is really confusing for a reader to understand why PDF are now considered and plotted.
Done. We revised the introduction part of this analysis: "*Figure 2 views the ΔT vs. d_{snow} relationship in a complementary form using the PDFs of ΔT for different snow depth regimes. This analysis allows a detailed evaluation of the snow regime-dependent ΔT separation by quantifying and comparing the modal value and width of the different conditional PDFs*."

P10 L1: "we divide the data into "shallow" (d_{snow}≤20cm) and "thick" (d_{snow}≥45cm) regimes". Why do you choose these thresholds to characterize shallow and deep snow?
Done. We explain that the Russian snow depth data show a non-normal distribution with a median value of 30 cm (SI Fig. 1). Therefore, we have chosen two classes, one with a threshold below (d_{snow}≤20cm) and one with a threshold above (d_{snow}≥45cm) of this median value. "*Since the Russian snow depths are clearly non-Normal in distribution (SI Fig. 1, with a median d_{snow} of 30 cm), we divide the data into "shallow" ($d_{snow} \leq 20$ cm) and "thick" ($d_{snow} \geq 45$cm) regimes to separate two snow depth regimes*."

P10 L3: "Based on the ∆T PDFs". There are lots of line of figure 2. Maybe a revision to figure 2 is necessary. Could you please be specific how these five models separate the regimes, while others do not?
Done. We agree that Fig. 2 was too busy. We improved the readability of Fig.2 by splitting it into two figures *Fig.2a and Fig.2b*. Fig.2a shows the PDFs for snow depth classes, while Fig.2b shows the PDFs for different air temperature regimes. We also use now the same colors for the different snow classes as in Fig. 4 to be consistent. The color of the different Tair classes is consistent in all relevant figures (Figs. 1, 2, 5). We improved the discussion of the separation of the PDFs: *"Based on the ΔT PDFs, five models (CoLM, CLM4.5, JULES, ORCHIDEE, MIROC-ESM) successfully separate the ΔT regimes under different snow depth conditions. Their simulated ΔT PDFs have a smaller modal value for thin snow than for thick snow, as in the observations. The other models clearly fail in separating the ΔT PDFs for the two different snow depth regimes."*

P10 L9-17: This is mainly a description of observations. I would suggest to fold in this paragraph into the text right after the introduction of figure 1.
Thanks, but we disagree because: First we discuss the general functional behavior of

ΔT vs. dsnow relationship, i.e. increasing ΔT with increasing dsnow (Fig.1). Here we start with observations, followed by model evaluation. This is then further supported and discussed by Fig.2a. Next we discuss then the impact of Tair on the ΔT vs. dsnow relationship (different colored curves in Fig. 1 and Fig.2b). Therefore, we keep this paragraph there at its place, because this is the introduction/motivating paragraph, why we expect and look at the impact of Tair on the the ΔT vs. dsnow relationship. To make this clearer, we add at the end of this paragraph "*Therefore, we can expect that the same thickness of snow in colder climates will provide greater insulation than it would in warmer climates.*"

P10 L19-20: "The observations in Figs. 1 and 2 indicate that snow under colder climates have greater insulation than under warmer climates." Please add an opening sentence to this paragraph. It would be great to give a reader a small hint about what is going to follow in the text below. This seems to be very typical for this manuscript, starting from the details and then reveal a motivation behind all of this.

Actually, we do have an opening paragraph above these lines, which gives the motivation looking at the impact of Tair on the the ΔT vs. dsnow relationship. There we explain the hypothesis that snow under colder climates have greater insulation than under warmer conditions, based on known facts: "*Both Figs. 1 and 2 further indicate that air-soil temperature differences are related to air temperature conditions. This is expected due to snow pack properties, particularly its density and moisture content, that affect the thermal conductivity of the snow. For example, the density of fresh fallen snow tends to be much lower under cold air temperatures than warm (Anderson, 1976), leading to increased insulation (larger ΔT). Snow densification is also a function of air temperature, for example, depth hoar metamorphosis of the snow pack, which produces more insulation (loosely packed depth-hoar crystals have very low thermal conductivity), is promoted by strong thermal gradients in the snow pack, and is typical of continental climates (e.g., Zhang et al., 1996). Therefore, we can expect that the same thickness of snow in colder climates will provide greater insulation than it would in warmer climates*." This paragraph gives the motivation and hypothesis. Then, with the next paragraph we discuss that indeed our observational analysis confirm this expected impact of Tair on the ΔT vs. dsnow relationship. According to this comments, we improved the connection between the two paragraphs. The first paragraph is followed by "*Indeed, our observational analysis (Figs. 1 and 2) confirm this. This is shown by a larger ΔT for colder T_{air} than for warmer T_{air} (for a certain snow depth) and a greater sensitivity of ΔT to changes in d_{snow} (Fig. 1), and by the larger modal value of the ΔT PDF for colder T_{air} than for warmer T_{air} (21 K for $T_{air} \leq -25$ ℃ and 9 K for -15 ℃ $< T_{air} \leq -5$ ℃; Fig. 2b). This is consistent with colder climates having lower density snow packs, and the differences are in line with measurements of snow density variability (Zhong et al., 2013).*"

P10 L32: "ΔT/ d_{snow}". On page 9, Line 25 there was a different notation for this relationship. I would suggest to make uniform notations. The sign "/" means

division and maybe there is a better choice, e.g. \Delta{T}(d_snow) P12 L16: "T_{soil} /T_{air}" Maybe it is possible to find a better notation. The sign / means division. Could you use T_obs(T_air)
Done. We agree with you. We changed all the "/" with "vs." throughout the paper.

P11 L5-7: "Our analysis (Fig. 1) indicates that some models (CLM4.5, CoLM, JULES) are better able to…". Please move this section of the text close to figure 1, the reader now needs to go back in text. Everything related to figure 1 needs to stay close to figure 1.
Thanks for the suggestion, but the structure as it stands seems clearer to us. First we discuss the general functional behavior of ΔT vs. dsnow relationship, i.e. increasing ΔT with increasing dsnow (Fig.1). Here we start with observations, followed by model evaluation. This is then further supported and discussed by Fig.2a. Next we discuss the impact of Tair on the ΔT vs. dsnow relationship (different colored curves in Fig. 1 and Fig.2b). Again, we start with observations, followed by model evaluation. Here we discuss in detail that part of Fig.1 which shows the snow-dependence on ΔT vs. dsnow relation (colored curves in Fig.1). To make this structure even more clear, we add one opening sentence before this model evaluation part starts. "*If we evaluate the models with respect to this observed impact of Tair to the ΔT vs. d_{snow} relationship, we demonstrate that some models (CLM4.5, CoLM, JULES) are better able to replicate the effect than others (LPJ-GUESS, MIROC-ESM, ORCHIDEE, UW-VIC) (Fig. 1)…"*

P11 L22: "strong". I also see lots of green and blue. I guess this is an indicator of weak correlation.
Sorry, we think this is a misunderstanding; we were not clear enough. Here "strong spatial variability" refers to the pronounced spatial variability in the correlation. We see in Fig.3 in some regions redish color (high correlation) or in other regions greenish color (low correlation). We revised this sentence "*The maps of the ΔT vs d_{snow} correlations in winter (Fig. 3) demonstrate a pronounced spatial variability in the ΔT vs d_{snow} relationship*." And, we re-ordered some text such that this sentence is then directly followed by the description of the details of regional variation in the correlation.

P11 L23: "but indicate that most models agree on the general large-scale pattern." Please describe what you see on the generated maps and only then state your conclusion. The first sentence is to describe what is going to happen in the rest of paragraph. It appears that the conclusions/details are always are beginning of many paragraphs. of course, there is a style to present a material this way, but it is only good for press-conferences when it is not enough time to describe assumptions, hypothesis, etc.
Done. We improved this paragraph to make it clearer. We start with the most obvious finding that the maps of the ΔT vs. dsnow correlations show a pronounced spatial variability in this correlation. Then, we describe the details of the observed spatial

variation. This is then followed by the model evaluation. Thus, we do have a clear and consistent presentation structure.

P12 L9: delete "Previous authors" and parentheses.
Done. We re-arranged this sentence. *"Previous studies have noted that the strength of relationship between T_{soil} and T_{air} is modulated by d_{snow} and the snow insulation effect increases only up to a limiting depth beyond which extra snow makes little difference to soil temperatures (Smith and Riseborough, 2002; Sokratov and Barry, 2002; Zhang, 2005; Lawrence and Slater, 2010)."*

P13 L18: "SI Fig.3". L24: "SI Fig.4". Bring this figure from the appendix to the text and describe how this figure is obtained. P15 L32: "SI Table 3". There is no reason to move table and figures to SI when they are cited so often. Ideally, it should be no mentioning of figure from the SI.
We moved SI Table 3 to the main text; it is Table 3 now. Both figures (SI Fig.3, SI Fig.4) are cited only once now and are supplementary information.

P13 L20-21: "Obvious outliers in the T_{soil}/T_{air} correlation maps (SI Fig. 3) are ISBA and UVic, which strongly overestimate the correlation (r > 0.9)". What does it physically mean?
Done. We improved this discussion. *"Obvious outliers in the T_{soil} vs T_{air} correlation maps are ISBA and UVic, which strongly overestimate the correlation (r > 0.9) over most of the Arctic. This indicates an underestimated snow insulation effect, and confirms the weak insulation in both models, which has been initially discussed based on the underestimated ΔT (Fig. 1) and weak correlation between ΔT and d_{snow} (Fig. 3)."*

P14 L10: "the biases range from -0.8 K to -4.7 K (SI Table 2)". Please state that this is for winter months in the text.
Done. *"The biases of winter mean air temperature ranges…"*

P14 L25-26: "underestimate" and "overestimate". Does it depend on the values of 250 kg/m3 used to convert the satellite measured SWE to snow depth.
No, this is not the case. Please see our answer to the related comment #3. Furthermore, we show in Table 3 (earlier SI Tab.3) that the model biases are consistent regardless of which observations are used for the model evaluation. The model biases with respect to the station data are consistent with those with respect to the GlobSnow data; the model ranking is not affected.

P14 L28-29: "The evaluation of the model performance for SWE compared to GlobSnow indicates the same bias characteristics as described here for snow depth (not shown)." This is really confusing... please re-write
Done. We deleted this sentence.

P15 L5-9: "Across-model differences in the interannual variability of winter precipitation do not translate simply to corresponding differences in the interannual d_{snow} variability (not shown)." This is another conclusion before the supporting statement. "For example, UVic calculates the (unrealistically) largest interannual dsnow variability in the boreal Europe permafrost region which is not reflected in the precipitation variability." This is the supporting statement.
This is a subjective criticism of style rather than substance. There are countless examples in the scientific literature of similar styles as we adopt here, that is a general point followed by a specific example. We do think this paragraph is clear.

P15 L13: "We have shown that the across-model spread in the representation of snow insulation effects". How did you show that? I think I lost something while reading Sections 3.1 and 3.2
This done in Figs.1-5 as well as in Table2.

P15 L15-16: "By considering the relationship plots and the conditional PDFs (Figs. 1, 2, 4, and 5)". Add figure #s here after "plots" and not all the figures in parentheses are related to PDFs.
Done. "*By considering the relationship plots (Figs. 1, 4 and 5), and the conditional PDFs (Fig. 2) we were able to classify the models…*"

P15 L17: "sort the models in terms of their snow insulation performance." Here "sort" means "classify"?
Done. Yes, this is what was meant. For clarity, we replace "sort" by "classify".

P15 L20-29: It is better to move this paragraph to the introduction, where the employed models are described.
Sorry, we do not agree with you. Here we describe those specific model characteristics and processes which explain why these models show a better performance than the others. It makes no sense to list which models have performed better in the introduction before the analysis is done.

P17 L1-2: "The results are further improved by updating the snow albedo and snow densification parameterization". References?
Done. "*Decharme et al. (2016) still showed that the ISBA results are further improved by updating the snow albedo and snow densification parameterization.*"

P17 L20-32: Move this paragraph into the above sections.
Sorry, we think this is a misunderstanding. This paragraph clearly belongs to the summary and conclusion. In this paragraph, we summarize the main findings.

P18 L1: "The primary aim of this study was to …". I thought that "the aim" is stated on line 14, page 17. Please clarify.
Done. We revised this sentence to "*This study uses the ensemble of models to*

document model performance with respect to T_{soil} versus T_{air} relationships..."

P18 L9: "Those models which show limited skill in snow insulation representation …have some deficiencies". Maybe over simplifications?
Done. We revised this to "*… have some deficiencies or oversimplifications...*"

P18 L24-30: The manuscript has not discussed the modeled area of permafrost prior to this paragraph. The conclusions are to summarize primary findings. Please move this paragraph into the main part of this manuscript.
Done. We agree with you and the other reviewer and we added an additional section "*5. Permafrost area*" in the results section in the paper with the additional new *Tab.4*, which presents the simulated permafrost area. "*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 km^2, which is comparable with the IPA map estimate (16.2 million km^2) (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 km^2) and higher (20.01 million km^2) 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 km^2). 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 km^2). 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."

P26 L3: "Table 1. PCN snow model details." Please specify how the upper 20cm of the soil column are resolved in each model. How is the heat capacity is calculated, may add another column.

According to our analysis, and as also noted by Referee #1, it is the snow layer in winter that is the key here, not the soil conditions (moisture, organic layer). Therefore, we think the current table contains enough information for the discussion in this paper, with its emphasis on snow above the soil. The general structure and some key parameterizations of snow processes can explain the main deficiencies in the simulated results. The interested reader may find such soil details in 3 papers we cite in the introduction on the model general characteristics.

P28 L4: "The dots represent the medians of…". Please use different symbols. If this is printed black&white or someone is color blind, it might be hard to differentiate the results. P29 L3: "Conditional probability density functions…". Why are they conditional? "color". Same comments as for the above figure. Please think how to make this plots available to color-blind people. P30 L3: "Spatial maps of the correlation…" Same comment as above P31 L4: "The dots represent the medians…" Same comment as above. Maybe use different symbols for dots. P32 L4: "The dots represent the medians…"

We do indeed sympathize with colour-blind readers. Personally journal should do more to encourage such consideration in their publications, but this is clearly an issue at the journal publisher level, not at the author or editor level. Our figure presentation is consistent with what is common in TC and other journals.

"Conditional probability" is a completely standard term in statistics. The "conditional" means the ΔT PDFs are PDFs for specific different conditions, namely for different conditions of air temperatures and snow depths.

Evaluation of air-soil temperature relationships simulated by land surface models during winter across the permafrost region

Wenli Wang[1+], Annette Rinke[1,2+], John C. Moore[1], Duoying Ji[1*], Xuefeng Cui[3], Shushi Peng[4,17,18], David M. Lawrence[5], A. David McGuire[6], Eleanor J. Burke[7], Xiaodong Chen[821], Bertrand Decharme[9]Christine Delire[9], Charles Koven[10], Andrew MacDougall[11], Kazuyuki Saito[12,15], Wenxin Zhang[13, 19], Ramdane Alkama[9,16], Theodore J. Bohn[8], Philippe Ciais[18], Christine Delire[9]Bertrand Decharme[9], Isabelle Gouttevin[4], Tomohiro Hajima[12], Gerhard Krinner[4,17], Dennis P. Lettenmaier[8], Paul A. Miller[13], Benjamin Smith[13], Tetsuo Sueyoshi[14], Artem B. Sherstiukov[20]

[1]College of Global Change and Earth System Science, Beijing Normal University, Beijing, China

[2]Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research (AWI), Potsdam, Germany

[3]School of System Science, Beijing Normal University, Beijing 100875, China

[4]CNRS, LGGE, 38041 Grenoble, France

[5]National Center for Atmospheric Research, Boulder, USA

[6]U.S. Geological Survey, Alaska Cooperative Fish and Wildlife Research Unit, University of Alaska Fairbanks, Fairbanks, AK, USA

[7]Met Office Hadley Centre, Exeter, UK

[8]School of Earth and Space Exploration, Arizona State University, Tempe, AZ, USA

[9]GAME, Unité mixte de recherche CNRS/Meteo-France, Toulouse cedex, France

[10]Lawrence Berkeley National Laboratory, Berkeley, CA, USA

[11]School of Earth and Ocean Sciences, University of Victoria, Victoria, BC, Canada

[12]Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan

[13]Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden

[14]National Institute of Polar Research, Tachikawa, Japan

[15]University of Alaska Fairbanks, Fairbanks, AK, USA

[16]l'Institute for Environment and Sustainability (IES), Ispra, Italy

[17]Université Grenoble Alpes, LGGE, 38041 Grenoble, France

[18]LSCE, CEA/CNRS/UVSQ, Saclay, France

[19]Center for Permafrost (CENPERM), Department of Geosciences and Natural Resource Management, University of Copenhagen, DK-1350 Copenhagen, Denmark

[20]All-Russian Research Institute of Hydrometeorological Information-World Data Centre, Obninsk, Russia

[21]Department of Civil and Environmental Engineering, University of Washington, Seattle, WA, USA

+ *these authors contributed equally to this work*

* *correspondence to*: Duoying Ji (duoyingji@bnu.edu.cn)

Abstract. A realistic simulation of snow cover and its thermal properties are important for accurate modelling of permafrost. We analyze simulated relationships between air and near-surface (20 cm) soil temperatures in the Northern Hemisphere permafrost region during winter, with a particular focus on snow insulation effects in nine land surface models and compare them with observations from 268 Russian stations. There are large across-model differences in the~~as expressed by~~ simulated differences between near-surface soil and air temperatures~~,~~ (*ΔT*)~~,~~ of 3 to 14 K, in the sensitivity of soil to air temperature~~gradients between soil and air temperatures~~ (0.13 to 0.96 ℃/℃), and in the relationship between *ΔT* and snow depth. The observed relationship between *ΔT* and snow depth can be used as a metric to evaluate the effects of each model's representation of snow insulation, and hence guide improvements to the model's conceptual structure and process parameterizations. Models with better performance apply multi-layer snow schemes and consider complex snow processes. Some models show poor performance in representing snow insulation due to underestimation of snow depth and/or overestimation of snow conductivity. Generally, models identified as most acceptable with respect to snow insulation simulate reasonable areas of near-surface permafrost (13.19 to 15.77~~12-16~~ million km^2). However, there is not a simple relationship between the sophistication~~quality~~ of the snow insulation in the acceptable models and the simulated area of Northern Hemisphere near-surface permafrost, ~~likely~~ because several other factors such as soil depth, the treatment of soil organic content and hydrology, and vegetation cover ~~differences in the treatment of soil organic matter, soil hydrology, surface energy calculations, and vegetation~~ also provide important controls on simulated permafrost distribution.

1 Introduction

Present-day permafrost simulations by global climate models are limited and future projections contain high, model-induced uncertainty (e.g., Slater and Lawrence, 2013; Koven et al., 2013). Most of the model biases and across-model differences in simulating permafrost area are due to biased atmospheric simulation e.g. of air temperature and -precipitation, biased simulation of snow and soil temperature, and the coupling between atmosphere and land-surface.~~simulations of the atmosphere (air temperature, precipitation), land (snow, soil temperature) and their coupling.~~ In winter, the snow insulation effect is a key process for the air-soil temperature coupling. Its strength depends on ~~both~~ the snow depth, areal coverage, snow density and conductivity~~quantity (depth, areal coverage) and snow quality (density, conductivity)~~ (see overview by Zhang, 2005). Many individual model studies have shown the strong impact of ~~a few~~ snow parameters on soil temperature simulations (e.g., recently, Langer et al., 2013; Dutra et al., 2012; Gouttevin et al., 2012; Essery et al., 2013; Wang et al., 2013; Jafarov et al., 2014). Most importantly, these studies showed that the consideration of wet snow metamorphism and snow compaction, improved snow thermal conductivity and multi-layer snow schemes can improve the simulation of snow dynamics and soil temperature. Parameterizations that take into account snow compaction (e.g. related to overburden pressure, thermal metamorphism and liquid water) work better than simpler schemes such as an exponential increase of density with time (Dutra et al., 2010). The influence of snow thermal conductivity on soil regime has been demonstrated by many model studies (e.g., Bartlett et al., 2006; Saha et al., 2006; Vavrus, 2007; Nicolsky et al., 2007; Dankers et al., 2011; Gouttevin et al., 2012). Winter soil temperature can change by up to 20 K simply by varying the snow thermal conductivity by 0.1-0.5 W m^{-1} K^{-1} (Cook et al., 2008). The snow insulation effect also plays an important role for the Arctic soil temperature response to climate change and therefore for future near-surface permafrost thawing and soil carbon vulnerability (e.g., Schuur et al., 2008). Shallower snow can reduce soil warming while shorter snow season can enhance soil warming (Lawrence and Slater, 2010). ~~(e.g., reduced soil warming due to shallower snow versus enhanced soil warming due to shorter snow season) (e.g., Lawrence and Slater, 2010), and therefore for future near-surface permafrost thawing and soil carbon vulnerability (e.g., Schuur et al., 2008).~~ The model skill in atmosphere-soil coupling with the concomitant snow cover in the Arctic is an important factor in the assessment of limitations and uncertainty of carbon mobility estimates (Schaefer et al., 2011).

The Snow Model Intercomparison Project (Snow MIP) (Essery et al., 2009) and the Project for Intercomparison of Land-Surface Parameterization Schemes (PILPS) Phase 2e (Slater et al., 2001) examined the snow simulations of an ensemble of land-surface schemes for the mid-latitudes. However, until now there has been no attempt to evaluate the air-soil temperature relationship in the Northern Hemisphere permafrost region and the detailed role of snow depth therein across an ensemble of models. In such an investigation, a first suitable approach is the evaluation of stand-alone (off-line) land surface models (LSMs). The retrospective (1960-2009) simulations from the model integration group of the Permafrost Carbon Network ("PCN"; http://www.permafrostcarbon.org) provide an opportunity to evaluate an ensemble of nine state-of-the-art LSMs. Here, the LSMs are run with observation-based atmospheric forcing, meaning that snow depth is not influenced by biases in the atmospheric forcing in a coupled model set-up. The evaluation of the offline modeled air temperature - snow depth - near-surface soil temperature relationship in winter is therefore important for revealing a model's skill in representing the effects of snow insulation.

Most of the LSMs participating in PCN are the land-surface modules of Earth System Models (ESMs) participating in the Coupled Model Intercomparison Project (CMIP5; http://cmip-pcmdi.llnl.gov/cmip5/) although in some cases different versions were used for PCN and CMIP5 simulations. Thus, the results we present can guide the corresponding evaluation of these ESMs, though analysis of coupled model results requires consideration of couplings between model components and is necessarily more complex.

The scope of the present study is to analyze the extent to which the ensemble of PCN models can reproduce the observed relationship between air and near-surface soil temperatures in the Northern Hemisphere permafrost region during winter, with a particular focus on the snow insulation effect. For the latter we analyze the impact of snow depth on the difference between near-surface soil and air temperatures. Our related key questions are: How well do the models represent the observed spatial pattern of the air-soil temperature difference in winter and its control by the snow depth? What is the range of the simulated air-soil temperature relationship across the model ensemble? How well do the models represent the observed magnitude and spatial pattern of the snow depth control over the air-soil temperature difference in winter? What is the range of these simulated relationships across the model ensemble? To the extent possible, we try to relate the performance level to the model's snow schemes. With this aim in mind, a simultaneous analysis of simulated air and near-surface soil temperatures, and snow depth is presented and compared with those from a novel set of Russian station observations. We focus here on a comprehensive Russian station data set because this has been compiled within PCN, and it is hard to find other station data sets which provide simultaneous observations of -both air and soil temperatures as well as snow depth over a long period.

In Sect. 2, we describe the model simulations, the station observations used for evaluation, and the analysis methods. In Sect. 3, we present a detailed analysis of near-surface air temperature - snow depth - soil temperature relationships in winter. In Sect. 4, we discuss the roles of atmospheric forcing and model processes. In Sect. 5, we investigate the relation of simulated snow insulation and permafrost area. We summarize our findings and present conclusions in Sect. 65.

2 Data and Analysis

2.1 Models

We use data from nine LSMs participating in the PCN, including CLM4.5, CoLM, ISBA, JULES, LPJ-GUESS, MIROC-ESM, ORCHIDEE, UVic, and UW-VIC. For detailed information about the models and simulations we refer to Rawlins et al. (2015) and, Peng et al. (2015), and Mc Guire et al. (2016). The total soil depth for soil thermal calculations ranges from 3 m (divided in 8 layers) in LPJ-GUESS to 250 m (divided into 14 layers) in UVic. The soil physical properties differ among the models as well, and four of them (CLM4.5, ISBA, UVic, UW-VIC) include organic horizons. Three models (ISBA, LPJ-GUESS, UW-VIC) did not archive soil sub-grid results and provide area-weighted ground temperature (i.e. averaged over wetlands and vegetated areas, and in some cases lake fractions).

Table 1 lists relevant snow model details. One model (UVic) uses an implicit snow scheme which replaces the upper soil column with snow-like properties, i.e. the near-surface soil layer takes the temperature of the air-snow interface. The other models use separate snow layers on top of the ground, either a single bucket (LPJ-GUESS, UW-VIC) or multi-layer snow schemes (CLM4.5, CoLM, ISBA, JULES, MIROC-ESM, ORCHIDEE). Snow insulation is explicitly considered in all models; increasing snow depth increases the insulation effect. Many models consider the effect of varying snow density on insulation (Table 1). This is parameterized by a snow conductivity-density relationship that describes how, as snow density increases, thermal conductivity increases, thereby reducing the snow insulation. Some of the models (LPJ-GUESS, MIROC-ESM, ORCHIDEE, UVic) use a fixed snow density, consider only dry snow and no compaction effects, while others represent liquid water in snow and different processes for snow densification such as mechanical compaction, and thermal and destructive metamorphism (Table 1).

The simulations were generally run for the period 1960-2009, although some simulations were stopped a few years earlier. Each model team was free to choose appropriate driving data sets for weather and climate, atmospheric CO_2, nitrogen deposition, disturbance, land cover, soil texture, etc. However, the climate forcing data (surface pressure, surface incident longwave and shortwave radiation, near-surface air temperature, wind and specific humidity, rain and snowfall rates) are from gridded observational datasets (e.g. CRUNCEP, WATCH) (SI Table 1). The exception is MIROC-ESM, which was run as a fully-coupled model, forced by its own simulated climate. Mean annual temperature of the MIROC-ESM simulations for the permafrost region were within the range (-7.2 to 2.2 ℃) of the other forcing data sets used in this study and the trend in near-surface air temperature (+0.03 ℃ yr^{-1}) was the same for all forcing data sets. However, MIROC-ESM had both the highest annual precipitation (range 433 to 686 mm) and the highest trend in annual precipitation (range -2.1 to +0.8 mm yr^{-1}) among the forcing data sets.

The spatial domain of interest is the Northern Hemisphere permafrost land regions. Our analysis is based on the 0.5° × 0.5° resolution gridded driving and modeled data for winter (DJF) 1980-2000.

2.2 Observations

A quality-checked data set of monthly near-surface air temperature, 20 cm soil temperatures and snow depth from Russian meteorological stations have been provided by the All-Russian Research Institute of Hydrometeorological Information-World Data Centre (RIHMI-WDC; http://meteo.ru/) ~~(Sherstyukov, 2008)~~. 579 stations report snow depth and 268 stations provide simultaneous data of all three variables. Ground surface temperature data are not available. A detailed description of dataset preparation is provided in Sherstiukov (2012a). Observing conditions at the Russian stations in all meteorological elements correspond with WMO standards. The observations presented have been included in data sets, such as GSOD, HadSRUT4 etc. and are widely used in climate research (e.g. Anisimov and Sherstiukov, 2016; Decharme et al. 2016; Park et al., 2014; Brun et al., 2013; Pavlov and Malkova, 2009; PaiMazumder et al., 2008). The soil temperature dataset was run through four independent methods of quality control (Sherstiukov, 2012b). However, some soil temperature observations could be disturbed by grass cutting during the warm season and the removal of organic materials, mainly at agricultural sites, which may affect the trend in warm season (Park et al., 2014), but this does not affect our results about the air - upper soil temperature relationship in winter.

Precipitation station data have been compiled from the Global Summary of the Day (GSOD) data set produced by the National Climatic Data Center (NCDC; http://www.ncdc.noaa.gov) for all of the stations that are included in the RIHMI-WDC data set. In addition to the station's ground snow depth observations w~~W~~e ~~also~~ use gridded snow water equivalent (SWE) data from the GlobSnow-2 product (http://www.globsnow.info/swe/), which has been produced using a combination of passive microwave radiometer and ground-based weather station data (Takala et al., 2011). ~~Snow depth was then calculated from SWE using a snow density of 250 kg m^{-3}.~~ Orographic complexity, vegetation cover, and snow state (e.g. wet snow) affect the accuracy of this product. When compared with ground measurements in Eurasia, the GlobSnow product ~~can~~ shows root-mean-square error (RMSE) values of 30 to 40 mm for SWE values below 150 mm, with retrieval uncertainty increases when SWE is above this threshold (e.g., Takala et al., 2011; Muskett, 2012; Klehemet et al., 2013).~~regional differences (of ca. 0.5-5 cm) with biases increasing with increasing SWE (e.g., Takala et al., 2011; Muskett, 2012).~~ To compare with station data, snow depth was then calculated from SWE using a snow density of 250 kg m^{-3}, which is a median observed value in winter. Zhong et al. (2013) report snow density values of 180-250 kg m^{-3} for tundra/taiga and 156-193 kg m^{-3} for alpine snow classes. Woo et al. (1983) report snow density values of 250-400 kg m^{-3} for various terrain types. Choice of density does not materially affect the results.

All these data have been compiled for winter (DJF) and the same time period of 1980-2000. This period was chosen because soil temperature data are sparse before 1980 and the JULES simulation stopped in the year 2000. Comparison of the simulations with the station data was done using a weighted bilinear interpolation from the 4 surrounding model grid points onto the station locations.

2.3 Analysis Methods

Our analysis is focused on the common winter (DJF) condition, although snow can begin in November and end at the beginning of May, but we checked that a different winter definition (NDJFMA) does not qualitatively~~substantially~~ change any of the inter-variables relationships found.the results. The focus in our study is on the evaluation of the simulated air-soil temperature relationships, modulated by snow depth. For this, we analyze the winter mean as well as the interannual variability (expressed as the standard deviation) of 4 key variables: near-surface air temperature (T_{air}), near-surface soil temperature (soil temperature at 20 cm depth; T_{soil}), snow depth (d_{snow}), and the difference between T_{soil} and T_{air}. This difference ΔT ($\Delta T = T_{soil} - T_{air}$) is called the air-soil temperature difference. By limiting our analysis to the winter only, we are able to attribute the across-model and model-to-observation differences in ΔT primarily to snow insulation effects. In winter, the effects of other factors (e.g. soil moisture, texture) on ΔT are much smaller than that of snow. Ground surface temperatures were not recorded in the Russian data set, but 20 cm soil depth temperatures were. To test how sensitive are results using 20 cm temperatures instead of ground surface, we also analyzed model simulated temperature differences between ground surface and T_{air}, and found no qualitative differences, hence justifying use of 20 cm observations.We assume that there is relatively little impact due to soil moisture and texture between surface and 20 cm depth in winter. Although we recognize the difference between ground surface and 20 cm soil temperatures and that soil organic layer could play a role in certain locations (e.g., Romanovsky and Osterkamp, 1995), ground surface temperatures are not recorded in the Russian data set, while 20 cm soil depth temperatures are, hence our choice. However, we find that the results do not significantly change when the model simulated temperature differences between ground surface and near-surface air temperature are used instead of between 20 cm soil and near-surface air temperatures.

We use the Pearson product-moment correlation coefficient and its significance (von Storch and Zwiers, 1999) to investigate the co-variability between ΔT and d_{snow} as well as between T_{soil} and its two forcing factors (T_{air} and d_{snow}). Before we compute the correlations we detrended the data by removing a least squares regression line. The calculated correlation maps (i.e. spatial distributions of correlation coefficients) based on model and observation data, allow the comparison of the spatial patterns of these relationships. We use correlation analysis to investigate the co-variability between ΔT and d_{snow} as well as between T_{soil} and its two forcing factors (T_{air} and d_{snow}). The input consists of detrended time series of winter means at each grid point. The calculated correlation maps (i.e. spatial distributions of correlation coefficients) based on model and observation data, allow the comparison of the spatial patterns of these relationships. Significance of correlation coefficients is estimated by the Student t-test.

To further examine the functional behavior between the key~~different~~ variables, we present relation diagrams between pairs of variables (e.g. variation of ΔT with change of d_{snow}). To evaluate the performance of the individual LSMs we calculate the ~~root-mean-square error (RMSE)~~RMSE between the observed and modeled relationships. We illustrate the dependence of ΔT vs. d_{snow} and T_{soil} vs. d_{snow} relations for three T_{air} ranges. To distinguish dry snow pack regimes from those where sporadic melt may occur even in winter, we split T_{air} into 3 regimes~~We illustrate the dependence on air temperatures by evaluating the models and observations for several different near-surface air temperature ranges. We split the data into 3 regimes~~: the coldest conditions ($T_{air} \leq$ -25 ℃) represent 24% of observations, the intermediate temperature conditions (-25 ℃ < $T_{air} \leq$ -15 ℃) represent 42% of the observations, and the warmest conditions (-15 ℃ < $T_{air} \leq$ -5 ℃) represent 34% of observations.~~ The principal motivation for such classifications is to distinguish dry snow pack regimes from those where sporadic melt may occur even in winter.~~ Hence it is an indirect separation of temperature-gradient metamorphosis regimes and density-gradient metamorphosis snow pack regimes. Additionally, we present conditional probability density functions (PDFs) of ΔT for different snow depth and air temperature regimes and compare the simulated PDFs with those obtained from station observations.

3 Results

3.1 Relationship between air – soil temperature difference and snow depth

The air-soil temperature difference (ΔT) - snow depth (d_{snow}) relationship in winter (Fig. 1) shows in the Russian station observations an increase of ΔT with increasing d_{snow}. The data exhibit a linear relation between ΔT and d_{snow} at relatively shallow snow depths with a trend towards asymptotic behavior at thicker snow, which is in agreement with earlier findings (Zhang, 2005; Ge and Gong, 2010; Morse et al., 2011). There is also significant scatter in the observation-based relationship indicated by the inter-quartile range in ΔT of 1.5-8.5 K at specific snow depth and air temperature regimes, likely resulting from complicating factors such as snow pack density and moisture content variability over the winter, as well as observational errors.

All models reproduce the observed relationship, i.e. increasing ΔT with increasing d_{snow}. However, Fig. 1 also shows a wide across-model spread in the simulated relationships, and that some of the models are not consistent with the behavior in the observations. Only three models (CLM4.5, CoLM, JULES) reproduce reasonably well the observed ΔT vs. d_{snow} relationship using a benchmark of RMSE < 5 K for all temperature regimes. In particular LPJ-GUESS, ORCHIDEE, UVic, UW-VIC, MIROC-ESM show large RMSE for cold air conditions. ISBA stands out overall, with a RMSE of 7-18 K in all temperature ranges. We conclude that these models do not adequately represent the features of the observed ΔT vs. d_{snow} relationship. The scatter in the modeled relationships, indicated by the inter-quartile range, is of the same order as in the observations, except for ISBA and MIROC-ESM which produce noticeably smaller variations.

Figure 2a views the ΔT vs. d_{snow} relationship in a complementary form using the PDFs of ΔT for different snow depth regimes. This analysis allows a detailed evaluation of the snow regime-dependent ΔT separation by quantifying and comparing the modal value and width of the different conditional PDFs. Since the Russian snow depths are clearly non-Normal in distribution (SI Fig. 1, with a median d_{snow} of 30 cm), we divide the data into "shallow" (d_{snow} ≤ 20 cm) and "thick" (d_{snow} ≥ 45cm) regimes to separate two snow depth regimes. The modal value of the station data ΔT PDF is 5 K for "shallow" snow and 14 K for "thick" snow - that is thick snow is a better insulator than thin snow. Based on the ΔT PDFs, five models (CoLM, CLM4.5, JULES, ORCHIDEE, MIROC-ESM) successfully separate the ΔT regimes under different snow depth conditions. Their simulated ΔT PDFs have a smaller modal value for thin snow than for thick snow, like in the observations. The other models clearly fail in separating the ΔT PDFs for the two different snow depth regimes. However, even for the five successful models, both the shapes and the modal values of the simulated PDFs differ from the observed PDF.

Both Figs. 1 and 2b further indicate that ΔT are related to T_{air} conditions. This is expected due to snow pack properties, particularly its density and moisture content, that affect the thermal conductivity of the snow. For example, the density of fresh fallen snow tends to be much lower under cold T_{air} than warm (Anderson, 1976), leading to increased insulation (larger ΔT). Snow densification is also a function of T_{air}, for example, depth hoar metamorphosis of the snow pack, which produces more insulation (loosely packed depth-hoar crystals have very low thermal conductivity), is promoted by strong thermal gradients in the snow pack, and is typical of continental climates (e.g., Zhang et al., 1996). Therefore, we can expect that the same thickness of snow in colder climates will provide greater insulation than it would in warmer climates.

Our analysis of observations (Figs. 1 and 2b) confirms i) a larger ΔT for colder T_{air} than for warmer T_{air} (for a given snow depth), ii) a greater sensitivity of ΔT to changes in d_{snow} (Fig. 1), and iii) larger modal value of the ΔT PDF for colder T_{air} than for warmer T_{air} (21 K for $T_{air} \leq$ -25 °C and 9 K for -15 °C < $T_{air} \leq$ -5 °C; Fig. 2b). These effects are consistent with colder climates having lower density snow packs, and the differences are in line with measurements of snow density variability (Zhong et al., 2013). Additionally, both the inter-quartile range in Fig. 1 and the width of the PDFs in Fig. 2b become larger as T_{air} cool. This may be related to the formation of depth hoar, which is a very good insulator and its varying presence in the snow pack decouples ΔT from d_{snow}. Cold, thin snow packs tend to contain much more low density depth hoar than warmer snow packs (e.g., Zhang et al., 1996; Singh et al., 2011). Continental regions have large annual temperature cycles, with greater interannual variability and thinner snow packs, than maritime ones. This variability leads to greater scatter and greater sensitivity of the ΔT vs. d_{snow} relationship in the cold winter regions. An additional cause of scatter is that the density of fresh-fallen snow decreases with falling temperature. Accordingly, we find in the cold T_{air} regime ($T_{air} \leq$ -25 °C) a larger ΔT in early winter (November-December) when the snow pack is composed of thin, low density fresh snow (and depth hoar) than in late winter (January-February) (SI Fig. 2). Under warm conditions (-15 °C < $T_{air} \leq$ -5 °C) such a separation is not observed.

If we evaluate the models with respect to this observed impact of T_{air} to the ΔT vs. d_{snow} relationship, we demonstrate that some models (CLM4.5, CoLM, JULES) are better able to replicate the effect than others (LPJ-GUESS, MIROC-ESM, ORCHIDEE, UW-VIC) (Fig. 1). The latter do not fully replicate the larger ΔT under cold T_{air} conditions. CLM4.5, CoLM and JULES capture a larger ΔT for colder T_{air} for a given d_{snow} in agreement with the observations. However, for shallow snow JULES simulates an increase of ΔT with increasing d_{snow} for all temperature ranges that is twice as large as observations. Two models (ISBA, UVic) clearly fail in this evaluation. Poor model performance in reflecting T_{air} influence on the ΔT vs. d_{snow} also manifests itself in regime separation of the PDFs (Fig. 2b). Some models do not separate the ΔT regimes under different T_{air} conditions well or at all (ISBA, LPJ-GUESS, MIROC-ESM, UVic), while others cannot capture the observed cold temperature regime features (i.e., too broad PDFs and shifts towards smaller modal values; ORCHIDEE, UW-VIC). The three models with reasonable inter-variable relations (CLM4.5, CoLM, JULES) also capture the regime separation in the PDFs. These three models as well as LPJ-GUESS and ORCHIDEE also represent the observed greater insulation of early winter snow packs under cold conditions (SI Fig. 2).

The maps of the *ΔT* vs. d_{snow} correlations in winter (Fig. 3) demonstrates a pronounced spatial variability in the *ΔT* vs. d_{snow} relationship. Highest positive correlation occurs in the region of the East Siberian Plain and Siberian High lands. In other regions, namely in Scandinavia, West Russian Arctic, West and Central Siberian Plains, the correlation is much weaker and often not statistically significant. These are the regions of large winter snow depth (Sect. 4.1.2) which are influenced by North Atlantic cyclonic activity which brings relatively warm moist air and heavy precipitation in winter (and a positive correlation between d_{snow} and T_{air}), leading to relatively small mean *ΔT*.

Some models (CLM4.5, CoLM, ORCHIDEE, UW-VIC) show a reasonable spatial pattern correlation coefficient (r ≥ 0.4) with observations, while the others do not (Fig. 3). Obvious outliers are the LPJ-GUESS and UVic models, which do not reproduce the observed pattern of correlation. UVic calculates a reverse spatial pattern comparing to that of the observations (e.g. significant positive correlation in West Siberian Plain and Central Siberian Highlands). LPJ-GUESS produces very few statistically significant correlations.
The air-soil temperature difference (*ΔT*) - snow depth (d_{snow}) relationship in winter (Fig. 1) shows that observations and all models produce a clear relationship between increase of *ΔT* and increases of d_{snow}. However, Fig. 1 also shows a wide across-model spread in the simulated relationships, and that some of the models are not consistent with the behavior in the observations. There is also significant scatter in the observation-based relationship, the inter-quartile range of *ΔT* is 1.5-8.5 K at specific snow depth and air temperature regimes, likely resulting from complicating factors such as snow pack density and moisture content variability over the winter, as well as observational errors. Similar ranges of variability are produced by several models (such as CLM4.5, CoLM and JULES), but other models (such as ISBA and MIROC-ESM) produce noticeably smaller variations.

The Russian station data and some model results exhibit a linear relation between winter *ΔT* and d_{snow} at relatively shallow snow depths with a trend towards asymptotic behavior at larger snow depths (Fig. 1), which is in agreement with earlier findings (Zhang, 2005; Ge and Gong, 2010; Morse et al., 2011). However, only three models (CLM4.5, CoLM, JULES) reproduce reasonably well the *ΔT* - d_{snow} relationship seen in the observational station data (Fig. 1) using a benchmark of RMSE < 5 K for all temperature regimes. In particular LPJ-GUESS, ORCHIDEE, UVic, UW-VIC, MIROC-ESM show large RMSE for cold air conditions. ISBA stands out overall, with a RMSE of 7-18 K in all temperature ranges. We conclude that these models do not adequately represent the features of the observed ΔT /d_{snow} relationship.

Figure 2 views the ΔT /d_{snow} relationship in the complementary form of the PDFs of ΔT for different snow depth and air temperature regimes. Since the Russian snow depths are clearly non-Normal in distribution (SI Fig. 1, with a median d_{snow} of 30 cm), we divide the data into "shallow" ($d_{snow} \leq 20$ cm) and "thick" ($d_{snow} \geq 45$cm) regimes. The modal value of the station data ΔT PDF is 5 K for "shallow" snow and 14 K for "thick" snow – that is thick snow is a better insulator than thin snow. Based on the ΔT PDFs, five models (CoLM, CLM4.5, JULES, ORCHIDEE, MIROC-ESM) successfully separate the ΔT regimes under different snow depth conditions, while the other models clearly fail for at least one of these snow depth regimes. However, even for the better models, both the shapes and the modal values of the simulated PDFs differ from the observed PDF.

Both Figs. 1 and 2 indicate that air-soil temperature differences are related to air temperature conditions. This is due to snow pack properties, particularly its density and moisture content, that affect the air-soil temperature difference. For example, the density of fresh fallen snow tends to be much lower under cold air temperatures than warm (Anderson, 1976), leading to increased insulation (larger ΔT). Snow densification is also a function of air temperature, for example, depth hoar metamorphosis of the snow pack, which produces more insulation (loosely packed depth-hoar crystals have very low thermal conductivity), is promoted by strong thermal gradients in the snow pack, and is typical of continental climates (e.g., Zhang et al., 1996).

The observations in Figs. 1 and 2 indicate that snow under colder climates have greater insulation than under warmer climates. This is shown by a larger ΔT for colder T_{air} than for warmer T_{air} (for a certain snow depth) and a greater sensitivity of ΔT to changes in d_{snow} (Fig. 1), and by the larger modal value of the ΔT PDF for colder T_{air} than for warmer T_{air} (21 K for $T_{air} \leq -25$ °C and 9 K for -15 °C $< T_{air} \leq -5$ °C; Fig. 2). This is consistent with colder climates having lower density snow packs, and the differences are in line with measurements of snow density variability (Zhong et al., 2013). Additionally, both the inter-quartile range in Fig. 1 and the width of the PDFs in Fig. 2 become larger as air temperatures cool. This may be related to the formation of depth hoar, which is a very good insulator and its varying presence in the snow pack decouples ΔT from d_{snow}. Cold, thin snow packs tend to contain much more low density depth hoar than warmer snow packs (e.g., Zhang et al., 1996; Singh et al., 2011). Continental regions have large annual temperature cycles, with greater interannual variability and thinner snow packs, than maritime ones. This variability leads to greater scatter and greater sensitivity of the ΔT /d_{snow} relationship in the cold winter regions. An additional cause of scatter is that the density of fresh-fallen snow decreases with falling temperature. Accordingly, we find in the cold air temperature regime ($T_{air} \leq -25$ ℃) a larger ΔT in early winter (November-December) when the snow pack is composed of thin, low density fresh snow (and depth hoar) than in late winter (January-February) (SI Fig. 2). Under warm conditions (-15 ℃ $< T_{air} \leq -5$ ℃) such a separation is not observed.

Our analysis (Fig. 1) indicates that some models (CLM4.5, CoLM, JULES) are better able to replicate the observed effect of air temperature on the ΔT /d_{snow} relationship than others (LPJ-GUESS, MIROC-ESM, ORCHIDEE, UW-VIC). The latter do not fully replicate the larger ΔT under cold air temperature conditions. CLM4.5, CoLM and JULES capture a larger ΔT for colder air temperatures for a given d_{snow} in agreement with the observations. However, for shallow snow JULES simulates twice as large increase of ΔT with increasing d_{snow} for all temperature ranges, compared with observations. Two models (ISBA, UVic) clearly fail in this evaluation. Poor model performance in reflecting air temperature influence on the ΔT /d_{snow} also manifests itself in regime separation of the PDFs (Fig. 2). Some models do not separate the ΔT regimes under different air temperature conditions well or at all (ISBA, LPJ-GUESS, MIROC-ESM, UVic), while others cannot capture the observed cold temperature regime features (i.e., too broad PDFs and shifts towards smaller modal values; ORCHIDEE, UW-VIC). The three models with reasonable inter-variable relations (CLM4.5, CoLM, JULES) also capture the regime separation in the PDFs. These three models as well as LPJ-GUESS and ORCHIDEE also represent the observed greater insulation of early winter snow packs under cold conditions (SI Fig. 2).

The maps of the ΔT /d_{snow} correlations in winter (Fig. 3) demonstrate the strong spatial variability in the ΔT /d_{snow} relationship, but indicate that most models agree on the general large-scale pattern. Some models (CLM4.5, CoLM, ORCHIDEE, UW-VIC) show a reasonable pattern correlation coefficient ($r \geq 0.4$) with observations, while the others do not. Most models simulate the highest positive correlation in the region of the East Siberian Plain and Siberian High lands. In some regions, namely in Scandinavia, West Russian Arctic, West and Central Siberian Plains, the correlation is much weaker and often not statistically significant. These are the regions of large winter snow depth (Sect. 4.1.2) which are influenced by North Atlantic cyclonic activity which brings relatively warm moist air and heavy precipitation in winter (and a positive correlation between snow depth and air temperature), leading to relatively small mean ΔT. Obvious outliers in the ΔT /d_{snow} correlation map are the LPJ GUESS and UVic models, which do not reproduce the observed pattern of correlation. UVic calculates a reverse pattern correlation than observations for many regions (e.g. significant positive correlation in West Siberian Plain and Central Siberian Highlands). LPJ-GUESS produces very few statistically significant correlations. The model correlations are likely highly sensitive to the quality of the snowfall forcing data, which is uncertain across much of the region due to limited station data that go into most global snowfall products (Hancock et al., 2014; Drobot et al., 2006).

3. 2 Variability of soil temperature with air temperature and snow depth

Next we assess whether or not the models can correctly reproduce the interannual near-surface soil temperature (T_{soil}) variability in relation to snow depth (d_{snow}) and near-surface air temperature (T_{air}) variability. Previous studies have notedPrevious authors (Smith and Riseborough, 2002; Sokratov and Barry, 2002; Zhang, 2005; Lawrence and Slater, 2010) have noted that the strength of relationship between T_{soil} and T_{air} is modulated by d_{snow} and the snow insulation effect increases only up to a limiting depth beyond which extra snow makes little difference to soil temperatures (Smith and Riseborough, 2002; Sokratov and Barry, 2002; Zhang, 2005; Lawrence and Slater, 2010). Zhang (2005) reportedreports that the limiting snow depth is approximately 40 cm.

To inspect the difference of the insulation effects on both sides of such a limiting snow depth, we investigate the T_{soil} vs. $T_{air}$$T_{soil}$ /T_{air} relationship under shallow ($d_{snow} \leq$ 20 cm) and thick ($d_{snow} \geq$ 45 cm) snow conditions. Observations showed that the slope of this relationship is higher when the snow cover is thin, compared with thicker snow conditions (e.g., for Yukon Territory in Canada; Karunaratne and Burn, 2003). Indeed, theOur Russian observation analysis (Fig. 4, Table 2) indicate a three times higher regression slope between T_{soil} and T_{air}stronger T_{soil} /T_{air} relationship (0.62 ℃/ ℃, R^2=0.8) under shallow snow pack than thicker snow conditions (0.21 ℃/ ℃, R^2=0.4). 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).

Figure 4 clearly shows that some models (CoLM, CLM45, JULES) can capture this modification of the T_{soil} vs. T_{air} 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 T_{soil} vs. T_{air} relationships from observations is smaller than 4 ℃. These models better reproduce the observed ΔT vs. d_{snow} relationship. Other models (LPJ-GUESS, MIROC-ESM, ORCHIDEE) strongly underestimate the increase of the T_{soil} vs. T_{air} 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 T_{soil} vs. T_{air} relation to snow conditions (Fig.4, Table 2). Another measure quantitatively confirms the same models behavior: The observed average d_{snow} 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 T_{air} 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 T_{soil} vs. T_{air} relation better than others (Table 2).~~The average d_{snow} 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 of 4.3 in the slopes of Fig. 4. The models that better reproduce the observed ΔT /d_{snow} relationship (CLM4.5, CoLM, JULES) reproduce the observed variation in the T_{soil}/T_{air} relation better than others.~~ 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 with unrealistic ΔT vs. d_{snow} relationships (ISBA, UVic) also fail in this evaluation of their T_{soil} vs. T_{air} relationship. They simulate a too strong sensitivity of T_{soil} to T_{air} ~~(~~(regression slopes larger~~The two models that had unrealistic ΔT /d_{snow} relationships (ISBA, UVic) also fail in this evaluation of their T_{soil} /T_{air} relationship. They simulate too strong T_{soil} /T_{air} relationships (gradients larger~~ than 0.9 ℃/ ℃, R^2>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 T_{soil} and T_{air} 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 slopestrong relationship shown in Fig. 4. The RMSE of their modeled T_{soil} vs. T_{air} T_{soil} /T_{air} relationships from observations reaches ca. 10 ℃ .

The T_{soil} vs. d_{snow} T_{soil} /d_{snow} relationship (Fig. 5) displays the variation of T_{soil} with changing snow depth and emphasizes the reduced sensitivity of T_{soil} to snow depth weakening role of snow depth for T_{soil} under thick snow conditions. With increasing d_{snow}, T_{soil} soil temperatures asymptotically converges towards a value of around 0 ℃. Overall, the Russian observations indicate that snow depth above about 80-90 cm has very little additional insulation effect on T_{soil}. Most of the models show consistent results with regard to this aspect, although the inter-quartile range of T_{soil} for specifica certain snow depths is quite large in some models (ISBA, ORCHIDEE, UVic, UW-VIC) (Fig. 5). The figure further points to the air temperature dependency of the relation. On average, for a given d_{snow}, a colder T_{soil} is observed for colder near-surface air temperatures, compared with warmer air temperatures. Most models can replicate this effect of air temperatures T_{air} on the T_{soil} vs. d_{snow} T_{soil} /d_{snow} relationship, though with differing accuracy. The RMSE between the observed and modeled relationships can reach ca. 10 ℃ and more (in ISBA, UVic, UW-VIC), particularly under cold conditions.

The spatial patterns of the correlation coefficients between T_{soil} and T_{air} (SI Fig. 3) and between T_{soil} and d_{snow} (SI Fig. 4) show a relatively large across-model scatter in the specificmany regions. Obvious outliers in the T_{soil} vs. T_{air} T_{soil} /T_{air} correlation maps (SI Fig. 3) are ISBA and UVic which strongly overestimate the correlation (r > 0.9) over most of the Arctic. This indicates an underestimated snow insulation effect, and confirms the weak insulation in both models, which we already discussed based on their underestimated ΔT (Fig. 1) and weak correlation between ΔT and d_{snow} (Fig. 3). Other models (LPJ-GUESS, ORCHIDEE, UW-VIC) also overestimate the correlation in some regions (e.g. western Russian Arctic, r > 0.7). Most of the simulated maps of T_{soil} vs. d_{snow} T_{soil} /d_{snow} correlation (SI Fig. 4) agree with the observations on a strong positive correlation in East Siberia. This is a region of relatively shallow snow (10-40 cm; Fig. 6) and there T_{soil} is very sensitive to variations in snow depth (e.g., Romanovsky et al., 2007). Comparing both simulated correlation maps, it is obvious that in this region, T_{soil} correlates more strongly with d_{snow} than with T_{air}, in agreement with the Russian data and earlier studies (Romanovsky et al., 2007; Sherstyukov, 2008).

4 Roles of atmospheric forcing and model processes

The across-model differences in the snow insulation effect, presented by the air temperature - snow depth - soil temperature relationships described above, are partially due to differences in the atmospheric forcing data and also due to differences in the snow and soil physics used in the LSMs. However, because the climate forcing data sets utilized with each model are observation-based (except for MIROC-ESM), obvious outliers in individual model performance likely mainly indicate poor or deficient physical descriptions of the air/snow/soil relations in that specific LSM.

4.1 Atmospheric forcing and snow depth

4.1.1 Air temperature and precipitation

Both near-surface air temperature (T_{air}) and precipitation are given by the climate forcing data sets (SI Table 1) for all models, except for MIROC-ESM which simulates both. The across-model differences in forcing T_{air}~~air temperature~~ used are relatively small and the simulated spatial patterns of temperature are very similar (SI Fig. 5). All forcing datasets are somewhat colder than Russian station data in their grid cells. The biases of winter mean T_{air} ranges from -0.8 K to -4.7 K (SI Table 2), reflecting biases in the climate forcing data used by the models. In contrast, MIROC-ESM has a positive (mean) T_{air}~~air temperatures~~ bias of +2.7 K.

The large-scale patterns of precipitation are similar across the models, but regional differences can be large (SI Fig. 6). The individual differences in winter precipitation range from~~-~~ -0.2 mm/day to +0.5 mm/day (SI Table 2) relative to the average of the Russian station data. Unfortunately, snowfall was archived in only a few models, however large-scale spatial patterns are similar across these models (SI Fig. 7).

4.1.2 Snow depth

The broad-scale spatial snow depth (d_{snow}) patterns are similar across the models and show general agreement with the observed patterns (Fig. 6). The well-pronounced areas of maximum winter d_{snow} (50-100 cm) are in Scandinavia, the Urals, the West Siberian Plain, Central Siberian Highlands, the Far East, Alaskan Rocky mountains, and Labrador Peninsula and isle of Newfoundland. However, large regional across-model variability is obvious. Some models (JULES, LPJ-GUESS, ORCHIDEE, UVic) underestimate d_{snow}, while others (CLM4.5, CoLM, ISBA, UW-VIC) overestimate it (Fig. 6; ~~SI~~ Table 3). The model biases are quite similar with respect to station observations and GlobSnow data. ~~The evaluation of the~~

model performance for SWE compared to GlobSnow indicates the same bias characteristics as described here for snow depth (not shown). It should be noted, that the models do not account for snowdrift. However, redistribution of snow due to wind is an important aspect, which makes comparison between in-situ measured and modeled snow depths difficult (e.g., Vionnet et al., 2013; Sturm and Stuefer, 2013; Gisnas et al., 2014).

Precipitation/snowfall across-model differences cannot be the primary explanation of these d_{snow} differences since some models (JULES, MIROC-ESM, ORCHIDEE) have positive bias in precipitation (> 0.2 mm/d, SI Table 2) but simulate much lower d_{snow} compared to other models (Fig. 6, SI Figs. 6, 7, SI Table 3). Across-model differences in the interannual variability of winter precipitation do not translate simply to corresponding differences in the interannual d_{snow} variability (not shown). For example, UVic calculates the (unrealistically) largest interannual d_{snow} variability in the boreal Europe permafrost region which is not reflected in the precipitation variability. These results indicate that the simulated snow depth is a function of both, the prescribed winter precipitation, and the model's snow energy and water balance.

4.2 Model processes

We have shown that the across-model spread in the representation of snow insulation effects (Sects. 3.1, 3.2) can not predominantly be explained by differences in the forcing data (Sect. 4.1), but to a large extent is due to the representation of snow processes in the models. By considering the relationship plots (Figs. 1, 4 and 5), and the conditional PDFs (Fig. 2) we were able to classifyBy considering the relationship plots and the conditional PDFs (Figs. 1, 2, 4, and 5) we were able to sort the models in terms of their snow insulation performance. In this section we discuss the influence of the different snow parameterizations in the models.

Models with better performance (CLM4.5, CoLM, JULES) apply multi-layer snow schemes. This allows them to simulate more realistic (stronger) insulation because they consider the snowpack's vertical structure and variability. They calculate the energy and mass balance in each snow layer, are able to capture nonlinear profiles of snow temperature, and can also account for thermal insulation within the snowpack such as when the upper layer thermally insulates the lower layers (e.g., Dutra et al., 2012). These models also incorporate storage and refreezing of liquid water within the snow, parameterize wet snow metamorphism, snow compaction, and snow thermal conductivity (Table 1), which have been found to be among the most important processes for good snow depth and surface soil temperature simulation (e.g., Wang et al., 2013).

An underestimated snow depth directly leads to insulation that is too weak in JULES, LPJ-GUESS, ORCHIDEE, and UVic (Fig. 6, ~~SI~~ Table 3). However only in ORCHIDEE and UVic does this lead to a significant underestimation of ΔT (~~SI~~ Table 3, SI Fig. 8) indicating bias compensation in the two other models. Thus, compensating error effects occur due to snow density and conductivity (SI Fig. 9, Table 1), which impact snow thermal insulation.

Our analysis showed that two models (ISBA, UVic) have T_{soil} vs. T_{air}~~T_{soil}/T_{air}~~ correlation that are too high indicating that they do not represent the modulation of the T_{soil} vs. T_{air}~~T_{soil}/T_{air}~~ relationship by snow depth (Fig. 4). This is consistent with their underestimation of ΔT (Figs. 1 and 2, SI Fig. 8, ~~SI~~ Table 3). In UVic, the snowpack is treated not as a separate layer but as an extension of the top soil layer and a combined surface-to-soil thermal conductivity is calculated (Table 1). Such a scheme largely negates or reduces the insulating capacity of snow (Slater et al., 2001). Koven et al. (2013) noted that such a scheme simulates very little warming of soil, and sometimes even cooling. The slightly underestimated snow depth (~~SI~~ Table 3, Fig. 6) contributes (but not as the primary factor) to reduced snow insulation, as reported for UVic (Avis, 2012).

ISBA strongly underestimates ΔT, while strongly overestimating d_{snow}, compared with observations (~~SI~~ Table 3, Fig. 6). However, ISBA uses the same atmospheric forcing data as JULES (accordingly the air temperature and precipitation are quite similar; SI Table 2). Also, the model's snow density (150-250 kg m^{-3}) is similar to other models (CLM45, CoLM, JULES) (SI Fig. 9) and in agreement with Zhong et al. (2013) who report snow density values of on 180-250 kg m^{-3} for tundra/taiga and 156-193 kg m^{-3} for alpine snow classes in winter. This apparent contradiction comes from the parameterization of snow cover fraction within each grid cell (SCF). The version of ISBA used here calculates a unique superficial soil temperature whether or not the soil is covered by snow and all the energy and radiative fluxes are area-weighted by SCF (equations 7 and 20 in *Douville et al.*, 1995). In order to get reasonable albedos in snow-covered forests, as is necessary when ISBA is coupled to the CNRM-CM climate model, the parameterization gives very low SCF in the boreal forest (between 0.2 and 0.5). Hence, snow insulates only 20% to 50% of the grid cell, despite fairly high snow depths. The heat fluxes from the snow-covered fraction are averaged with the fluxes from the snow-free surface, strongly concealing the actual insulating effect of snow and underestimating it over the grid cell. Using the detailed snow model Crocus (Brun et al., 1992; Vionnet et al., 2012) with a SCF equal to 100% leads to an almost perfect simulation of near-surface soil temperature over Northern Eurasia (Brun et al., 2013). A similar experiment with ISBA and a SCF equal to 100% (Decharme et al., ~~2015~~2016) leads to good performances showing that the low ΔT in ISBA despite high snow depth in the present study is mostly due to this sub-grid snow fraction. Decharme et al. (2016) still showed that the ISBA ~~The~~ results are further improved by updating the snow albedo and snow densification parameterization.

Interestingly, the ORCHIDEE performance in simulating snow depth and ΔT is similar to UVic (underestimation of d_{snow} and ΔT; ~~SI~~ Table 3). However, ORCHIDEE can better represent the observed T_{soil} vs. T_{air}~~T_{soil}/T_{air}~~ relationship and its modulation due to snow pack. ORCHIDEE employs, similarly to UVic, a fixed snow density and thermal conductivity. However, in contrast with UVic, ORCHIDEE applies a multi-layer scheme and simulates heat diffusion in the snowpack in up to 7 discrete layers (Table 1; Koven et al., 2009). This helps resolving the snow thermal gradients between the top and the base of the snow cover, and might explain how some of the snow insulation effects are reasonably represented in ORCHIDEE, despite the simpler treatment of temperature diffusion.

5 Permafrost area

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 km^2, which is comparable with the IPA map estimate (16.2 million km^2) (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 km^2) and higher (20.01 million km^2) 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 km[2]). 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).

6 Summary and conclusions

The aim of this work was to evaluate how state-of-the-art LSMs capture the observed relationship between winter near-surface soil and air temperatures (T_{soil}, T_{air}) and their modulation by snow depth (d_{snow}) and climate regime. We presented some benchmarks to evaluate model performance. The results are based on the comparison of LSMs with a novel comprehensive Russian station data set. The presented relation diagrams of T_{soil} and the difference of T_{soil}-T_{air} 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. The results are based on the comparison of LSMs with a comprehensive Russian station data set.

We see large differences across the models in their mean air-soil temperature difference (ΔT) of 3 to 14 K, in the sensitivity of near-surface soil temperature to and air temperature (T_{soil} vs. T_{air}) in the gradient between near-surface soil and air temperatures (T_{soil} /T_{air}) (0.49 to 0.96 ℃/℃ for shallow snow, 0.13 to 0.93 ℃/℃ for thick snow), and in the increase of ΔT with increasing snow depth (modal value of ΔT PDF: 0 to 10 K for shallow snow, 5 to 21 K for thick snow). Most of the nine models compare to the observations reasonably well (observations: ΔT = 12 K, modal ΔT values of 5 K for shallow snow and of 14 K for thick snow, T_{soil} vs. T_{air} T_{soil} /T_{air} = 0.62 ℃/℃ for shallow snow, T_{soil} vs. T_{air} T_{soil} /T_{air}= 0.21 ℃/℃ for thick snow). Several models also capture the modulation by air temperature condition (larger increase in ΔT with increasing d_{snow} under colder conditions) and display the control of snow depth on T_{soil} (weaker T_{soil} vs. T_{air} T_{soil} /T_{air} relationship under thicker snow). However, while they generally capture these observed relationships, their strength can differ in the individual models. Two models (ISBA, UVic) show the largest deficits in snow insulation effects and cannot separate the ΔT regimes neither for different snow depths nor for different air temperature conditions.

This study uses the~~The primary aim of this study was to use this~~ ensemble of models to document model performance with respect to T_{soil} versus T_{air} relationships, and to identify those with better performance, rather than to quantify the best model. We were able to attribute performance strength/weakness to snow model features and complexity. Models with better performance apply multi-layer snow schemes and consider complex snow processes (e.g. storage and refreezing of liquid water within the snow, wet snow metamorphism, snow compaction). Those models which show limited skill in snow insulation representation (underestimated ΔT, very weak dependency of ΔT on d_{snow}, almost unity ratio of T_{soil} vs. T_{air}~~T_{soil} /T_{air}~~) have some deficiencies or over simplification in the simulation of heat transfer in snow and soil layer, particularly in the representation of snow depth and density (conductivity). We also emphasize that compensating errors in snow depth and conductivity can occur. For example, an excessive correlation between T_{soil} and T_{air} can be attributed to excessively high thermal conductivity even when the snow depth is correctly (or over) simulated. This finding underscores the need for detailed model evaluations using multiple, independent performance metrics to establish that the models get the right functionality for the right reason. It should be noted that the treatment of ground properties, particularly soil organic matter and soil moisture/ice content, also affect the simulated winter ground temperatures. The specific evaluation of these individual processes is more robustly investigated with experiments conducted for individual models (e.g. recently, Wang et al., 2013; Gubler et al., 2013; Decharme et al., 2015).

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 km^2). 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.
~~A realistic simulation of the snow is a key pre-requisite for accurate modeling of the soil thermal dynamics across the permafrost region. The areal cover of Northern Hemisphere near-surface permafrost varies greatly across the nine models in the hindcast simulation (1960-2009). Some of the better performing snow insulation effect models (CLM4.5, JULES)~~

~~simulate a near-surface permafrost area of 12 to 16 million km^2, which is comparable with the IPA map estimate (16 million km^2) (Brown et al., 1997; Slater and Lawrence, 2013). CoLM and ORCHIDEE, identified as reasonable models with respect to snow insulation, simulate much lower (7 million km^2) and higher (20 million km^2) areas, respectively. However, ISBA, one of the two models that showed rather limited skill in representing snow insulation effects, also simulates the highest permafrost area (20 million km^2). This is consistent with previous studies (e.g., Vavrus, 2007; Koven et al., 2013) which concluded that first-order control on modeled near-surface permafrost distribution is the representation of the air-to-surface soil temperature difference. When the models with poor snow models are eliminated, there is no clear relationship between the quality of the snow insulation in the models and the simulated area of permafrost, likely because several other factors such as differences in the treatment of soil organic matter, soil hydrology, surface energy calculations, model soil column depth, and vegetation also provide important controls on simulated permafrost distribution (Marchenko and Etzelmüller, 2013).~~

Acknowledgments. The data will be made available through the National Snow and Ice Data Center (NSIDC; http://nsidc.org); the contact person is Kevin Schaefer (kevin.schaefer@nsidc.org). This study was supported by the Permafrost Carbon Vulnerability Research Coordination Network, which is funded by the U.S. National Science Foundation (NSF). Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. E.J.B. was supported by the Joint UK DECC / Defra Met Office Hadley Centre Climate Program (GA01101). E.J.B., S.P., P.C. and G.K. were supported by the European Union Seventh Framework Program (FP7/2007-2013) under grant agreement n°282700. T.J.B. was supported by grant 1216037 from the NSF Science, Engineering and Education for Sustainability (SEES) Post-Doctoral Fellowship program. B.D., R.A. and C.D. were supported by the French Agence Nationale de la Recherche under agreement ANR-10-CEPL-012-03. This research was sponsored by the Integrated approaches and impacts, China Global Change Program (973 Project), National Basic Research Program of China Grant 2015CB953602 and the National Natural Science Foundation of China Grant 40905047.

References

Anderson, E.A.: A point energy and mass balance model of a snow cover, Office of Hydrology, National Weather Service, Silver Spring, Maryland, NOAA Technical Report NWS 19, 1976.

Andreadis, K., Storck, P. and Lettenmaier, D.P.: Modeling snow accumulation and ablation processes in forested environments, Water Resour. Res., 45, W05429, doi:10.1029/2008WR007042, 2009.

Anisimov, O.A., Sherstiukov A.B. Evaluating the effect of environmental factors on permafrost factors in Russia, Earth's Cryosphere, XX(2), 90-99, 2016.

Avis, C.A.: Simulating the present-day and future distribution of permafrost in the UVic Earth System Climate Model, Dissertation, University of Victoria, Canada, 274pp, 2012.

Bartlett, P.A., MacKay, M.D., Verseghy, D.L.: Modified snow algorithms in the Canadian Land Surface Scheme: model runs and sensitivity analysis at three boreal forest stands, Atmosphere-Ocean, 44, 207–222, 2006.

Best, M.J., and 16 co-authors: The Joint UK Land Environment Simulator (JULES), model description—Part 1: energy and water fluxes, Geosci. Model. Dev., 4, 677–699, doi:10.5194/gmd-4-677-2011, 2011.

Boone, A., and Etchevers, P.: An intercomparison of three snow schemes of varying complexity coupled to the same land-surface model: Local scale evaluation at an Alpine site, J. Hydrometeor., 2, 374-394, 2001.

Brown, J., Ferrians, O.J., Heginbottom, J.A. and Melnikov, S.E.: International Permafrost Association Circum-Arctic Map of Permafrost and Ground Ice Conditions, scale 1:10,000,000, Circum-Pacific Map Series, USGS Circum-Pacific Map Series, Map CP-45, 1997.

Brun, E., David, P., Sudul, M. and Brunot, G.: A numerical model to simulate snow cover stratigraphy for operational avalanche forecasting, J. Glaciol., 38, 13–22, 1992.

Brun, E., Vionnet, V., Boone, A., Decharme, B., Peings, Y., Valette, R., Karbou, F. and Morin, S.: Simulation of northern Eurasian local snow depth, mass and density using a detailed snowpack model and meteorological reanalysis, J. Hydrometeorol., 14, 203–214, doi:10.1175/jhm-d-12-012.1, 2013.

Cook, B.I., Bonan, G.B., Levis, S. and Epstein, H.E.: The thermoinsulation effect of snow cover within a climate model, Clim. Dyn., 31, 107-124, doi:10.1007/s00382-007-0341-y, 2008.

Dai, Y., Zeng, X., Dickinson, R.E., Baker, I., Bonan, G.B., Bosilovich, M.G., Denning, A.S., Dirmeyer, P.A., Houser, P.R., Niu, G., Oleson, K.W., Schlosser, C.A. and Yang, Z.: The

Common Land Model (CLM), Bull. Am. Meteorol. Soc., 84, 1013–1023, doi:10.1175/BAMS-84-8-1013, 2003.

Dankers, R., Burke, E.J., and Price, J.: Simulation of permafrost and seasonal thaw depth in the JULES land surface scheme, The Cryosphere, 5, 773-790, doi:10.5194/tc-5-773-2011, 2011.

Decharme, B., Brun, E., Boone, A., Delire, C., Le Moigne, P. and Morin, S.: Impacts of snow and organic soils parameterization on North-Eurasian soil temperature profiles simulated by the ISBA land surface model, The Cryosphere, 10, 853–877, doi: 10.5194/tc-10-853-2016, 2016.

Decharme, B., Brun, E., Boone, A., Delire, C., Le Moigne, P. and Morin, S.: Impacts of snow and organic soils parameterization on North-Eurasian soil temperature profiles simulated by the ISBA land surface model, The Cryosphere Discussions, 9, 6733-6790, doi:10.5194/tcd-9-6733-2015, 2015.

Douville, H., Royer, J.-F. and Mahfouf, J.-F.: A new snow parameterization for the Meteo-France climate model. Part 1: Validation in stand-alone experiments, Clim. Dyn., 12, 21–35, 1995.

Drobot, S., Maslanik, J., Herzfeld, U.C., Fowler, C. and Wu, W.: Uncertainty in temperature and precipitation datasets over terrestrial regions of the Western Arctic, Earth Interactions, 10 (23), 1-17, 2006.

Dutra, E., Viterbo, P., Miranda, P.M.A. and Balsamo, G.: Complexity of snow schemes in a climate model and its impact on surface energy and hydrology, J. Hydrometeorol., 13, 521–538, doi:10.1175/jhm-d-11-072.1, 2012.

Dutra, E., Balsamo, G., Viterbo, P., Miranda, P.M.A., Beljaars, A., Schär, C. and Elder, K.: An improved snow scheme for the ECMWF land surface model: description and offline validation, J. Hydrometeorol., 11, 899-916, 2010.

Essery, R., Morin, S., Lejeune, Y. and Ménard, C.B.: A comparison of 1701 snow models using observations from an alpine site, Adv. Water Resour., 55, 131–148, doi:10.1016/j.advwatres.2012.07.013, 2013.

Essery, R.L.H, Rutter, N., Pomeroy, J., Baxter, R., Staehli, M., Gustafsson, D., Barr, A., Bartlett, P. and Elder, K.: SnowMIP2: An evaluation of forest snow process simulations, Bull. Am. Meteorol. Soc., 90, 1120-1135, doi:10.1175/2009BAMS2629.1, 2009.

Ge, Y. and Gong, G.: Land surface insulation response to snow depth variability, J. Geophys. Res., 115, D08107, doi:10.1029/2009JD012798, 2010.

Gerten, D., Schaphoff, S., Haberlandt, U., Lucht, W. and Sitch, S.: Terrestrial vegetation and water balance: Hydrological evaluation of a dynamic global vegetation model, J. Hydrol., 286, 249–270, 2004.

Gisnas, K., Westermann, S., Schuler, T., Litherland, T., Isaksen, K., Boike, J. and Etzelmuller, B.: A statistical approach to represent small-scale variability of permafrost temperatures due to snow cover, The Cryosphere, 8, 2063-2074, doi: 10.5194/tc-8-2063-2014, 2014.

Gouttevin, I., Menegoz, M., Domine, F., Krinner, G., Koven, C.D., Ciais, P., Tarnocai, C. and Boike, J.: How the insulating properties of snow affect soil carbon distribution in the continental pan-Arctic area, J. Geophys. Res., 117, G02020, doi:10.1029/2011JG001916, 2012.

Gubler, S., Endrizzi, S., Gruber, S. and Purves, R.S.: Sensitivities and uncertainties of modeled ground temperatures in mountain environments, Geosci. Model Dev., 6, 1319-1336, doi:10.5194/gmd-6-1319-2013, 2013.

~~Hancock, S., Huntley, B., Ellis, R. and Baxter, R.: Biases in reanalysis snowfall found by comparing the Jules land surface model to GlobSnow, J. Clim., 27(2), 624-632, 2014.~~

Jafarov, E.E., Nicolsky, D.J., Romanovsky, V.E., Walsh, J.E., Panda, S.K. and Serreze, M.C.: The effect of snow: How to better model ground surface temperatures, Cold Regions Sci. Technol., 102, 63-77, doi:10.1016/j.coldregions.2014.02.007, 2014.

Ji, D., Wang, L., Feng, J., Wu, Q., Cheng, H., et al.: Description and basic evaluation of Beijing Normal University Earth System Model (BNU-ESM) version 1, Geosci. Model Dev., 7, 2039-2064, 2014.

Jordan, R.: A one-dimensional temperature model for a snow cover, technical documentation for SNTHERM.89, Special Report 91-16, U.S. Army Cold Regions Research and Engineering Laboratory, Hanover, N.H, 1991.

Karunaratne, K.C., and Burn, C.R.: Freezing n-factors in discontinuous permafrost terrain, Takhini River, Yukon Territory, Canada Proc. 8th Int. Conf. on Permafrost, Zurich, Eds. M. Phillips, S.M. Springman and L.U.Arenson, pp 519–24, 2003.

Klehmet, K., Geyer, B., and Rockel, B.: A regional climate model hindcast for Siberia: analysis of snow water equivalent, The Cryosphere, 7, 1017-1034, doi:10.5194/tc-7-1017-2013, 2013.

Koven, C.D., Riley, W.J. and Stern, A.: Analysis of Permafrost Thermal Dynamics and Response to Climate Change in the CMIP5 Earth System Models, J. Clim., 26, 1877–1900. doi:10.1175/JCLI-D-12-00228.1, 2013.

Koven, C., Friedlingstein, P., Ciais, P., Khvorostyanov, D., Krinner, G. and Tarnocai, C.: On the formation of high-latitude soil carbon stocks: Effects of cryoturbation and insulation by organic matter in a land surface model, Geophys. Res. Lett., 36, L21501, doi:10.1029/2009GL040150, 2009.

Langer, M., Westermann, S., Heikenfeld, M., Dorn, W., Boike, J.: Satellite-based modeling of permafrost temperatures in a tundra lowland landscape, Remote Sensing of Environment, 135 , pp. 12-24, doi:10.1016/j.rse.2013.03.011, 2013.

Lawrence, D.M., and Slater, A.G.: The contribution of snow condition trends to future ground climate, Clim. Dyn., 34, 969-981, doi:10.1007/s00382-009-0537-4, 2010.

Ling, F., and Zhang, T.: Sensitivity study of tundra snow density on surface energy fluxes and ground thermal regime in Northernmost Alaska, Cold Regions Sci. Technol., 44, 121–130, 2006.

Marchenko, S. and Etzelmüller, B.: Permafrost: Formation and Distribution, Thermal and Mechanical Properties. In: John F. Shroder (ed.) Treatise on Geomorphology, Volume 8, pp. 202-222. San Diego: Academic Press, 2013.

McGuire, A.D., et al.: A model-based analysis of the vulnerability of carbon in the permafrost region between 1960 and 2009, Global Biogeochemical Cycles, in press, 2016

Meissner, K.J., Weaver, A.J., Matthews, H.D. and Cox, P.M.: The role of land-surface dynamics in glacial inception: A study with the UVic earth system model, Clim. Dyn., 21, 515-537, 2003.

Morse, P.D., Burn, C.R., and Kokelj, S.V.: Influence of snow on near-surface ground temperatures in upland and alluvial environments of the outer Mackenzie Delta, Northwest Territories, Can. J. Earth Sci., 49, 895–913, doi:10.1139/E2012-012, 2012.

Muskett, R.: Remote sensing, model-derived and ground measurements of snow water equivalent and snow density in Alaska, Int. J. Geosci., 3, 1127-1136, doi:10.4236/ijg.2012.35114, 2012.

Nicolsky, D.J., Romanovsky, V.E., Alexeev, V.A. and Lawrence, D.M.: Improved modelling of permafrost dynamics in a GCM land-surface scheme, Geophys. Res. Lett., 34, L08501, doi:10.1029/2007GL029525, 2007.

Oleson, K.W., Lawrence, D.M., Bonan, G.B., Drewniak, B., Huang, M., Koven, C.D., Levis, S., Li, F., Riley, W.J., Subin, Z.M., Swenson, S.C., Thornton, P.E., Bozbiyik, A., Fisher, R., Kluzek, E., Lamarque, ,J.-F., Lawrence, P.J., Leung, L.R., Lipscomb, W., Muszala, S., Ricciuto, D.M., Sacks, W., Sun, Y., Tang, J., Yang, Z.-L.: Technical description of version

4.5 of the Community Land Model (CLM). NCAR Technical Note NCAR/TN-503+STR, doi: 10.5065/D6RR1W7M, 2013.

PaiMazumder, D., Miller, J., Li, Z., Walsh, J. E., Etringer, A., McCreight, J., Zhang, T., Mölders, N. Evaluation of Community Climate System Model soil temperatures using observations from Russia, Theoretical and Applied Climatology , 94(3),187-213, 2008.

Paquin, J.-P., and Sushama, L.: On the Arctic near-surface permafrost and climate sensitivities to soil and snow model formulations in climate models, Clim. Dyn., 44, 203-228, doi:10.1007/s00382-014-2185-6, 2015.

Park, H., Fedorov, A.N., Zheleznyak, M.N., Konstantinov, P.Y. and Walsh, J.E.: Effect of snow cover on pan-Arctic permafrost thermal regimes, Clim. Dyn., 44, 2873-2895, doi:10.1007/s00382-014-2356-5, 2015.

Park, H., A.B. Sherstiukov, A.N. Fedorov, I.V. Polyakov, and J.E Walsh: An observation-based assessment of the influences of air temperature and snow depth on soil temperature in Russia, Environ. Res. Lett. 9, doi:10.1088/1748-9326/9/6/064026, 2014.

Pavlov, A.V., Malkova, G.V. Small-scale mapping of trends of the contemporary ground temperature changes in the Russian North, Earth's Cryosphere, XIII(4), 32-39, 2009.

Peng, S., Ciais, P., Krinner, G., Wang, T., Gouttevin, I., McGuire, A., Lawrence, D., Burke, E., Chen, X., Decharme, B., Koven, C., MacDougall, A., Rinke, A., Saito, K., Zhang, W., Alkama, R., Bohn, T.J., Delire, C., Hajima, T., Ji, D., Lettenmaier, D.P., Miller, P.A., Moore, J.C., Smith, B. and Sueyoshi, T.: Simulated high-latitude soil thermal dynamics during the past four decades, The Cryosphere, 10, 1–14, doi:10.5194/tc-10-1-2016, 2015.

Rawlins, M., McGuire, A., Kimball, J., Dass, P., Lawrence, D., Burke, E., Chen, X., Delire, C., Koven, C., MacDougall, A., Peng, S., Rinke, A., Saito, K., Zhang, W., Alkama, R., Bohn, T.J., Ciais, P., Decharme, B., Gouttevin, I., Hajima, T., Ji, D., Krinner, G., Lettenmaier, D.P., Miller, P.A., Moore, J.C., Smith, B. and Sueyoshi, T.: Assessment of model estimates of land–atmosphere CO 2 exchange across Northern Eurasia, Biogeosciences, 12, 4385-4405, doi:10.5194/bg-12-4385-2015, 2015.

Riseborough, D.W.: An analytical model of the ground surface temperature under snow cover with soil freezing, 58th Eastern Snow Conference, Ottawa, Ontario, Canada, 2001

Romanovsky, V.E., Sazonova, T.S., Balobaev, V.T., Shender, N.I. and Sergueev, D.O.: Past and recent changes in air and permafrost temperatures in eastern Siberia, Glob. Planet. Change, 56, 399-413, doi:10.1016/j.gloplacha.2006.07.022, 2007.

Romanovsky, V.E., and Osterkamp, T.E.: Interannual variations of the thermal regime of the active layer and near-surface permafrost in Northern Alaska, Permafrost Periglac. Process., 6, 313–335, 1995.

Saha, S., Rinke, A., Dethloff, K. and Kuhry, P.: Influence of complex land surface scheme on Arctic climate simulations, J. Geophys. Res., 111, D22104, doi:10.1029/2006JD007188, 2006.

Schaefer, K., Zhang, T., Bruhwiler, L. and Barrett, A.P.: Amount and timing of permafrost carbon release in response to climate warming, Tellus B, 63, 165–180, doi:10.1111/j.1600-0889.2011.00527.x, 2011.

Schuur, E.A.G., and 18 coauthors: Vulnerability of permafrost carbon to climate change: Implications for the global carbon cycle, Bioscience, 58, 701-714. doi:10.1641/b580807, 2008.

Sherstyukov, A.B.: Correlation of soil temperature with air temperature and snow cover depth in Russia, Earth's Cryosphere, 12, 79-87, 2008.

Sherstiukov, A. Dataset of daily soil temperature up to 320 cm depth based on meteorological stations of Russian Federation, RIHMI-WDC, 176, 224-232, 2012a.

Sherstiukov, A. Statistical quality control of soil temperature dataset, RIHMI-WDC, 176, 224-232, 2012b.

Singh, V.P., Singh, P. and Haritashya, U.: Encyclopedia of snow, ice and glaciers, Springer, 1240p, 2011.

Slater, A.G., Schlosser, C.A. and Desborough, C.E.: The representation of snow in land-surface schemes: Results from PILPS 2(d), J. Hydrometeorol., 2, 7-25, 2001.

Slater, A., and Lawrence, D.: Diagnosing present and future permafrost from climate models, J. Clim., doi:10.1175/JCLI-D-12-00341.1, 2013.

Smith, M.W., and Riseborough, D.W.: Climate and the limits of permafrost: a zonal analysis, Permafrost Periglac. Process., 13, 1-15, doi: 10.1002/ppp.410, 2002.

Sokratov, S.A., and Barry, R.G.: Intraseasonal variation in the thermoinsulation effect of snow cover on soil temperatures and energy balance, J. Geophys. Res., 107, doi:10.1029/2001JD000489, 2002.

Sturm, M., Holmgren, J., König, M. and Morrris, K.: The thermal conductivity of seasonal snow, J. Glaciol., 43 (143), 26–41, 1997.

Sturm, M., and Stuefer, S.: Wind-blown flux rates derived from drifts at arctic snow fences, J. Glaciol., 59 (213), 21-34, 2013.

Swenson, S.C., and Lawrence, D.M.: A new fractional snow-covered area parameterization for the Community Land Model and its effect on the surface energy balance, J. Geophys. Res., 117, D21107, doi:10.1029/2012JD018178, 2012.

Takala, M., Luojus, K., Pulliainen, J., Derksen, C., Lemmetyinen, J., Karna, J.P., Koskinen, J. and Bojkov, B.: Estimating Northern Hemisphere snow water equivalent for climate research through assimilation of space-borne radiometer data and ground-based measurements, Remote Sens. Environ., 115, 3517-3529, 2011.

Takata, K., Emori, S. and Watanabe, T.: Development of the minimal advanced treatments of surface interaction and runoff, Glob. Planet. Change,38, 209–222, 2003.

Vavrus, S.J.: The role of terrestrial snow cover in the climate system, Clim. Dyn., 20, 73-88, doi:10.1007/s00382-007-0226-0, 2007.

Vionnet, V., Brun, E., Morin, S., Boone, A., Faroux, S., Moigne, P.L., Martin, E. and Willemet, J.-M.: The detailed snowpack scheme Crocus and its implementation in SURFEX v7.2, Geosci. Model Dev., 5, 773–791, 2012.

Vionnet, V., Guyomarch, G., Martin, E., Durand, Y., Bellot, H., Bel, C. and Puglièse, P.: Occurrence of blowing snow events at an alpine site over a 10-year period: observations and modelling, Adv. Water Resour., 55, 53-63, 2013.

Von Storch, H. and Zwiers, F.W.: Statistical Analysis in Climate Research, Cambridge University Press, Cambridge, 484pp., 1999.

Wang, T., Ottle, C., Boone, A., Ciais, P., Brun, E., Morin, S., Krinner, G., Piao, S. and Peng, S.: Evaluation of an improved intermediate complexity snow scheme in the ORCHIDEE land surface model, J. Geophys. Res., 118, 6064-6079, doi:10.1002/jgrd.50395, 2013.

Wania, R., Ross, I. and Prentice, I.C.: Integrating peatlands and permafrost into a dynamic global vegetation model: 2. Evaluation and sensitivity of vegetation and carbon cycle processes, Global Biogeochem. Cycles, 23, GB3015, doi:10.1029/2008GB003413, 2009.

Woo, M., Heron, R., Marsh, P. and Steer, P.: Comparison of weather station snowfall with winter snow accumulation in high arctic basins, Atmosph-Ocean, 21 (3), 312-325, doi:10.1080/07055900.1983.9649171, 1983.

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

Zhang, T., Osterkamp, T.E. and Stamnes, K.: Influence of the depth hoar layer of the seasonal snow cover on the ground thermal regime, Water Resour. Res., 32, 2075–2086, doi:10.1029/96WR00996, 1996.

Zhong, X., Zhang, T. and Wang, K.: Snow density climatology across the former USSR, The Cryosphere, 8, 785-799, doi:10.5194/tc-8-785-2014, 2013.

Tables

Table 1. PCN snow model details.

Model **Reference for snow scheme**	**Snow scheme**[1]	**Snow layers**	**Water phases**	**Liquid water treatment**[2]	**Snow density**[3]	**Snow thermal conductivity**[4]
CLM4.5 Swenson and Lawrence, 2012 Oleson et al., 2013	ML	Dynamic (max. 5)	Liquid, Ice	Bucket-type prognostic in each layer	depends on snow depth; compaction [3) a,b,c]	quadratic equation on ρ
CoLM Dai et al., 2003 Ji et al. 2014	ML	Dynamic (max. 5)	Liquid, Ice	Bucket-type prognostic in each layer	depends on snow depth; compaction [3) a,b,c]	quadratic equation on ρ
ISBA Boone and Etchevers, 2001	ML	Static (3)	Liquid, Ice, Vapor	Diagnosed from snow temperature, mass, density	compaction [3) a,b]	quadratic equation on ρ, contribution due to vapor transfer
JULES Best et al., 2011	ML	Dynamic (max. 3)	Liquid, Ice, Vapor	Bucket-type prognostic in each layer	compaction [3) a]	power equation on ρ
LPJ-GUESS Gerten et al., 2004 Wania et al., 2009	BL	Static (1)	Ice	Not represented	fixed 362 $kg\ m^{-3}$	fixed 0.196 $Wm^{-1}K^{-1}$
MIROC-ESM Takata et al., 2003	ML	Dynamic (max. 3)	Ice	Not represented	fixed 300 $kg\ m^{-3}$	fixed 0.3 $Wm^{-1}K^{-1}$
ORCHIDEE Gouttevin et al.,2012	ML	Dynamic (max. 7)	Ice	Not represented	fixed 330 $kg\ m^{-3}$	fixed 0.25 $Wm^{-1}K^{-1}$ for tundra, 0.042 $Wm^{-1}K^{-1}$ for taiga
UVic Meissner et al., 2003 Avis, 2012	I	Static (1)	Ice	Not represented	fixed 330 $kg\ m^{-3}$	bulk conductivity
UW-VIC Andreadis et al., 2009	BL	Dynamic (max. 2)	Liquid, Ice, Vapor	Constant liquid water holding capacity	compaction [3) a,b]	fixed 0.7 $Wm^{-1}K^{-1}$

[1] ML: Multi-layer, BL: Bulk-layer, I: Implicit; according to Slater et al. (2001)

[2] Not represented means dry snow

[3] Processes for densification of the snow: a) mechanical compaction (due to the weight of the overburden), b) thermal metamorphosis (via the melting–refreezing process), c) destructive metamorphism (crystal breakdown due to wind, thermodynamic stress); Anderson (1976), Jordan (1991), Kojima (1967)

[4] quadratic equation on ρ according to Jordan (1991), Anderson (1976); contribution due to vapor transfer according to Sun et al.(1999)

Table 2. Sensitivity of near-surface soil temperature (T_{soil}) to air temperature (T_{air}) in winter (DJF) calculated by the slopes of the linear regression between T_{soil} (℃) and T_{air} (℃) for different regimes of snow depth (d_{snow}), using data from all Russian station grid points and 21 individual winter 1980-2000. All relationships are statistically significant at p ≤ 0.01.

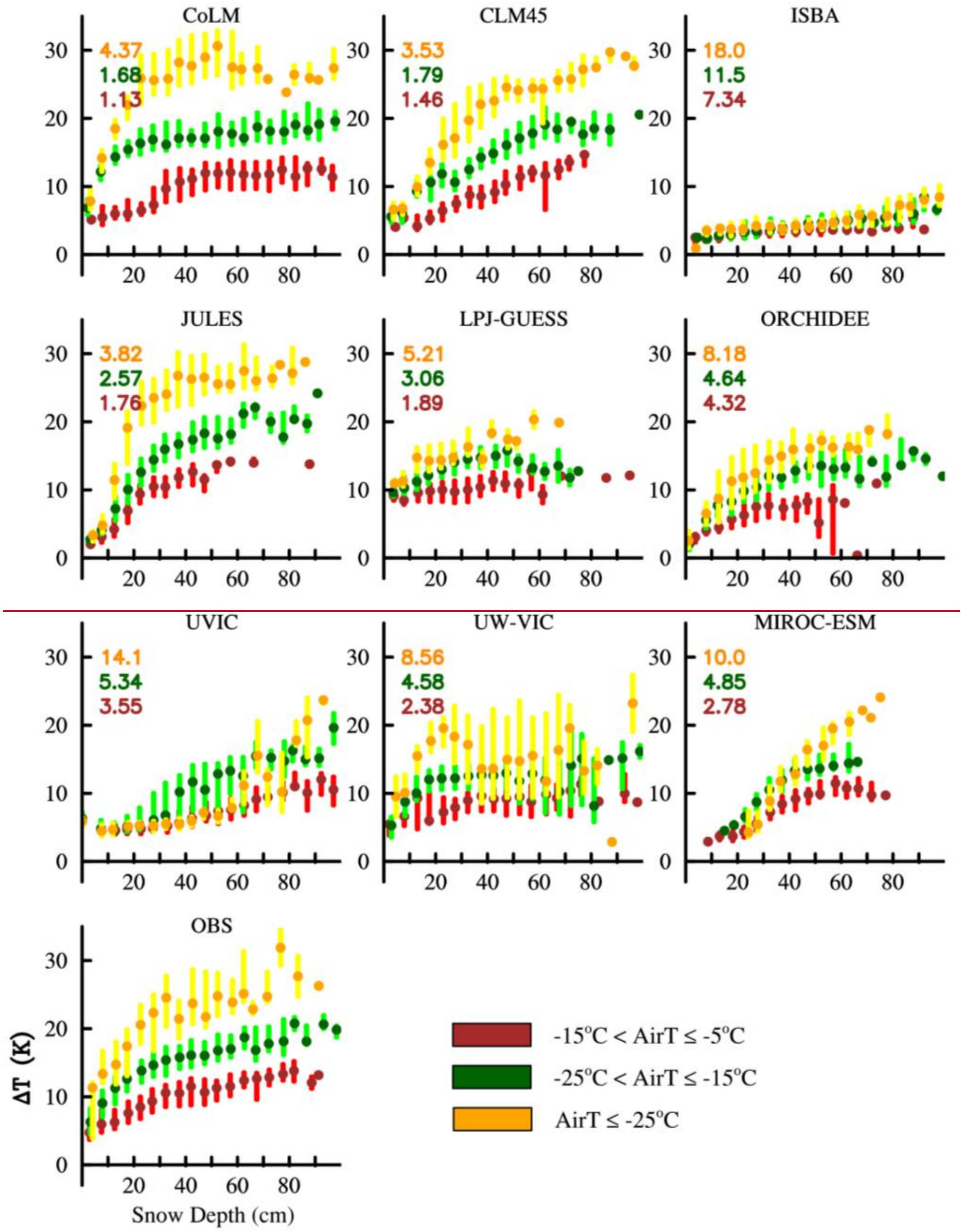
	Snow depth regimes			
	Shallow $d_{snow} \leq 20$ cm		Thick $d_{snow} \geq 45$ cm	
	T_{soil} vs. T_{air} ~~T_{soil}/T_{air}~~ (℃/℃)	R^2	T_{soil} vs. T_{air} ~~T_{soil}/T_{air}~~ (℃/℃)	R^2
Observation	0.62	0.79	0.21	0.41
CLM4.5	0.69	0.89	0.33	0.56
CoLM	0.49	0.73	0.13	0.44
ISBA	0.93	0.98	0.93	0.94
JULES	0.68	0.77	0.19	0.46
LPJ-GUESS	0.73	0.89	0.52	0.75
MIROC-ESM	0.78	0.98	0.49	0.67
ORCHIDEE	0.86	0.83	0.56	0.64
UVic	0.96	0.97	0.81	0.68
UW-VIC	0.54	0.74	0.76	0.65

Table 3. Russian-station-location averaged error statistics for snow depth (cm) and temperature difference between 20 cm soil and air temperature (ΔT; K) for winter 1980-2000. For each variable, the maximum available number of observations (n) is used. $Mean^{St,GS}$ and $stdev^{St,GS}$ are the observed mean and interannual variability (standard deviation), while stdev is the standard deviations of each model. Bias is the mean error 'simulation minus observation' and rmse is the root-mean-square error. The statistics for snow depth is given based on both station observation (St) and GlobSnow (GS) data.

	Snow depth (n=579) meanSt= 26.4 cm, mean^{GS}=23.4 cm stdevSt= 9.0 cm, stdev^{GS}= 6.5 cm					ΔT (n=268) meanSt= 11.9 K stdevSt= 2.3 K		
	biasSt	rmseSt	bias^{GS}	rmse^{GS}	stdev	biasSt	rmseSt	stdev
CLM4.5	11.5	18.1	14.3	18.1	5.8	2.3	4.1	2.2
CoLM	15.6	21.4	17.8	22.1	9.8	2.7	3.7	2.4
ISBA	13.0	18.8	15.7	19.8	9.5	-8.4	9.1	0.9
JULES	-4.1	14.1	-1.3	12.8	7.7	-0.8	4.2	3.2
LPJ-GUESS	-5.3	17.3	-2.5	16.0	5.0	-0.7	3.7	1.7
MIROC-ESM	-0.4	17.9	1.9	14.0	6.3	-4.9	6.7	2.0
ORCHIDEE	-8.7	16.5	-5.3	15.3	6.9	-5.2	6.0	1.9
UVic	-3.7	18.9	-0.5	16.8	9.4	-5.1	6.5	1.4
UW-VIC	12.5	19.8	15.0	20.0	10.4	-1.3	4.8	2.1

Table 4. Permafrost area, defined as maximum seasonal active layer thickness < 3 m in 1960 (Mc Guire et al., 2016). The IPA map estimate is 16 million km^2 (Brown et al., 1997; Slater and Lawrence, 2013).

Land Surface Model	Snow Insulation skill	Permafrost Area (10^6 km^2)
CLM4.5	High	15.77
CoLM	High	7.62
ISBA	Low	20.86
JULES	High	13.19
LPJ-GUESS	Medium	17.41
MIROC-ESM	Medium	13.02
ORCHIDEE	Medium	20.01
UVic	Low	16.47
UW-VIC	Medium	17.56


CoLM
4.37
1.68
1.13
CLM45
3.53
1.79
1.46
ISBA
18.0
11.5
7.34
JULES
3.82
2.57
1.76
LPJ-GUESS
5.21
3.06
1.89
ORCHIDEE
8.18
4.64
4.32
UVIC
14.1
5.34
3.55
UW-VIC
8.56
4.58
2.38
MIROC-ESM
10.0
4.85
2.78
OBS
ΔT (K)
Snow Depth (cm)
-15°C < AirT ≤ -5°C
-25°C < AirT ≤ -15°C
AirT ≤ -25°C

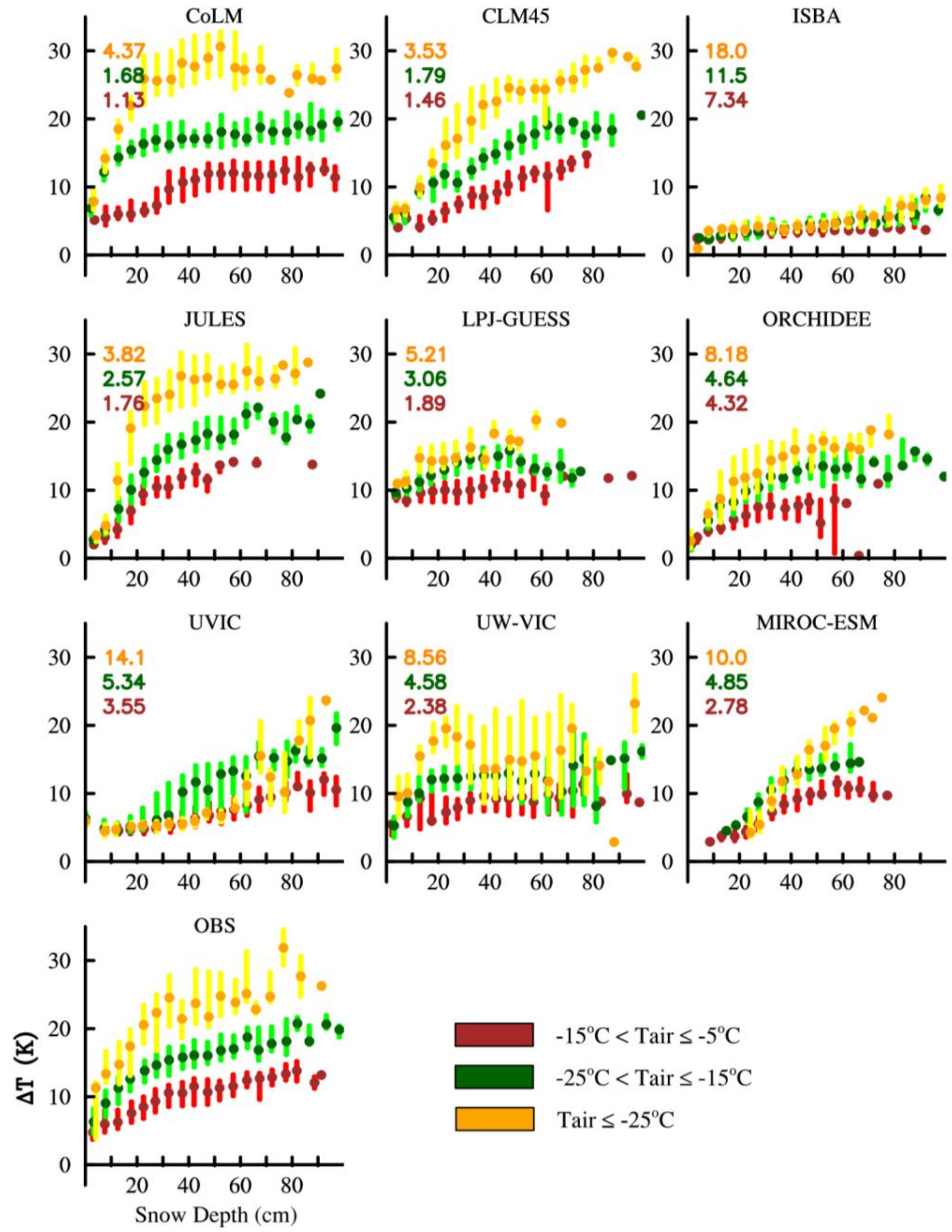
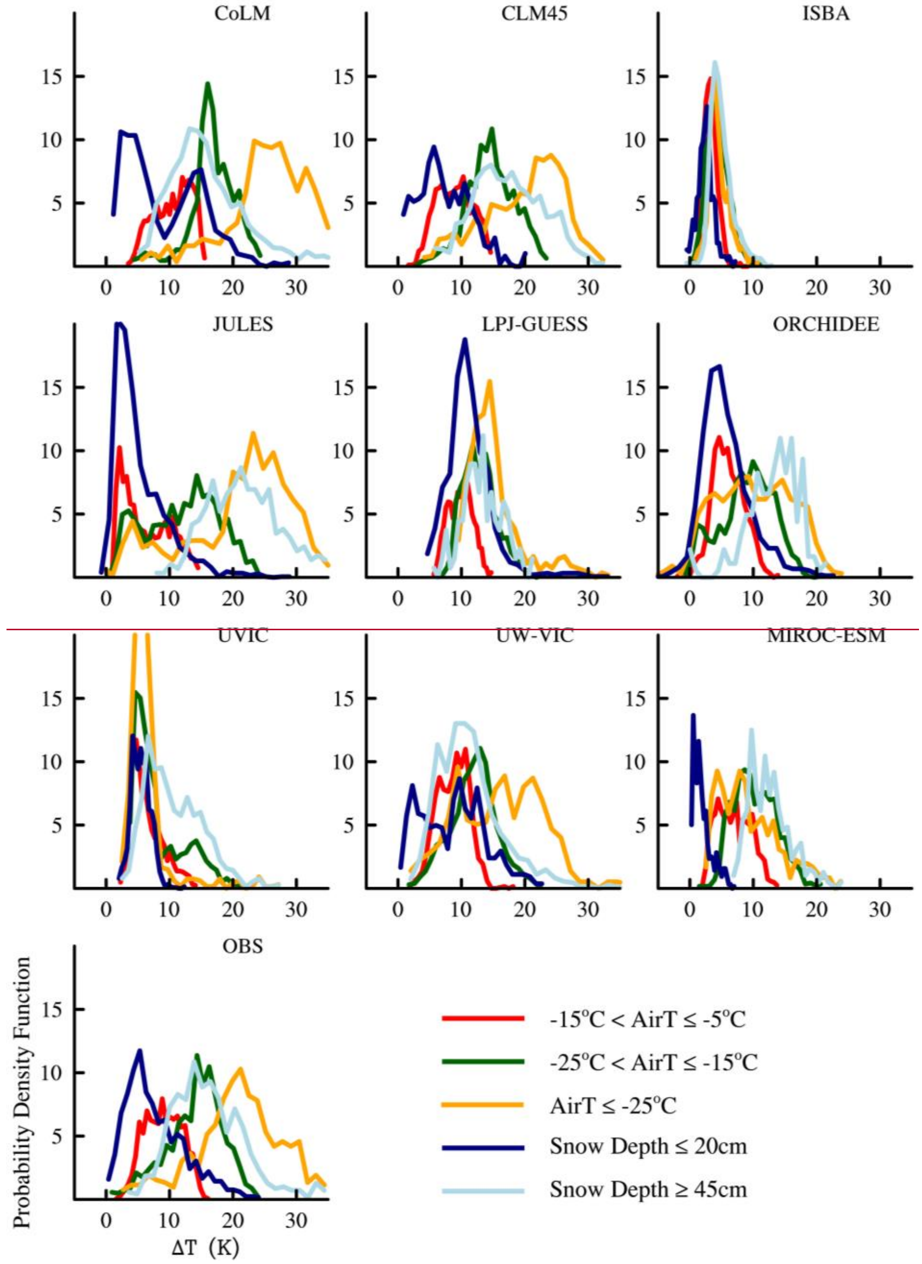
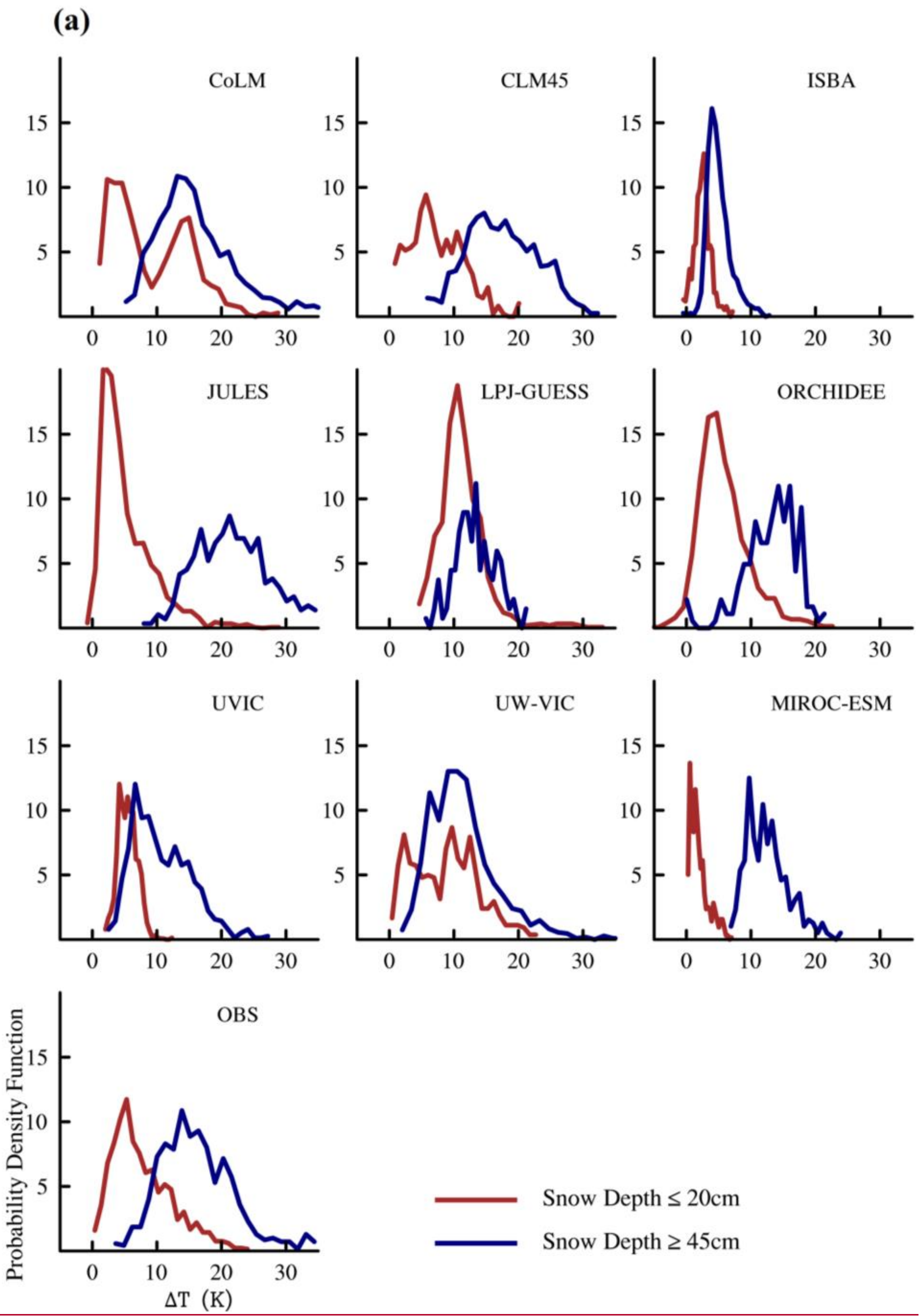


Figure 1. Variation of ΔT (K), the difference between soil temperature at 20 cm depth and air temperature) with snow depth (cm) for winter 1980-2000. The dots represent the medians of 5 cm snow depth bins and the upper and lower bars indicate the 25th and 75th percentiles, calculated from all Russian station grid points (n=268) and 21 individual winters. The numbers in each model panel indicate the RMSE between the observed and modeled relationship. Color represents different air temperature regimes.

1 
CoLM
CLM45
ISBA
JULES
LPJ-GUESS
ORCHIDEE
UVIC
UW-VIC
MIROC-ESM
OBS
Probability Density Function
ΔT (K)
-15°C < AirT ≤ -5°C
-25°C < AirT ≤ -15°C
AirT ≤ -25°C
Snow Depth ≤ 20cm
Snow Depth ≥ 45cm


(a)
CoLM
CLM45
ISBA
JULES
LPJ-GUESS
ORCHIDEE
UVIC
UW-VIC
MIROC-ESM
OBS
Probability Density Function
ΔT (K)
Snow Depth ≤ 20cm
Snow Depth ≥ 45cm

1

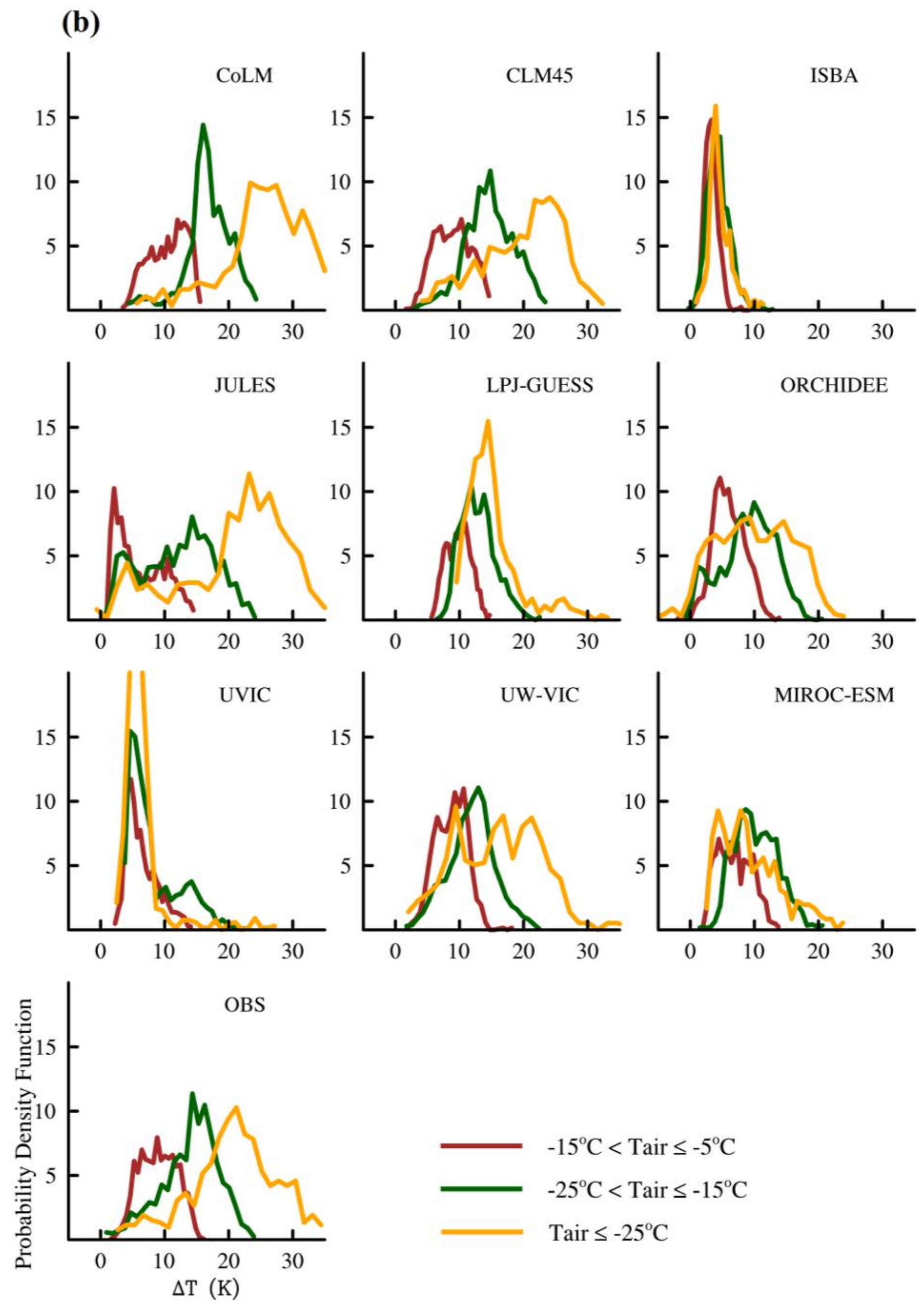


Figure 2. Conditional probability density functions (PDFs) of ΔT (K), the difference between soil temperature at 20 cm depth and air temperature for (a) different snow depth classes and (b) air temperature regimes, for winter 1980-2000.for different snow depth and air temperature regimes (color) for winter 1980-2000.

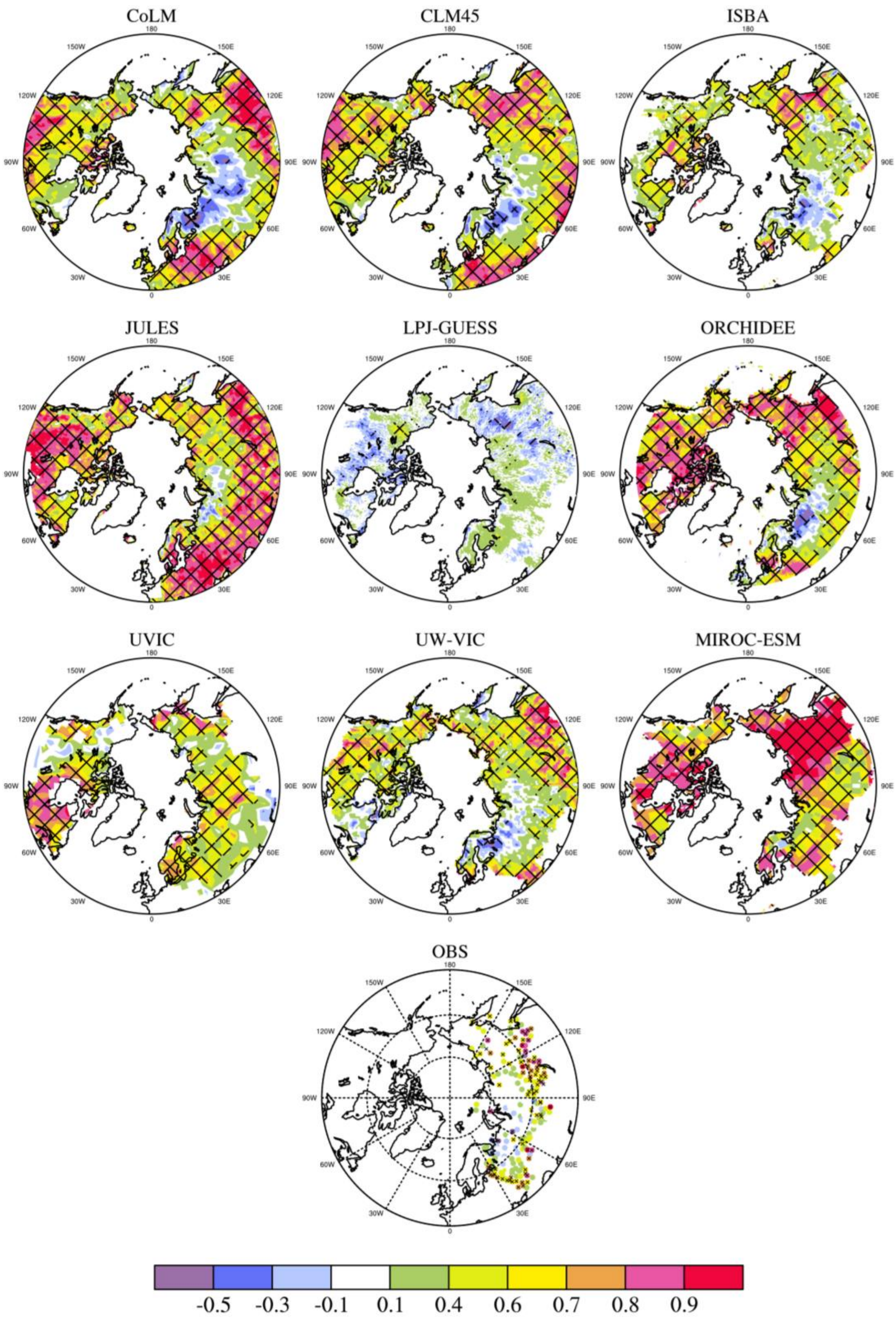
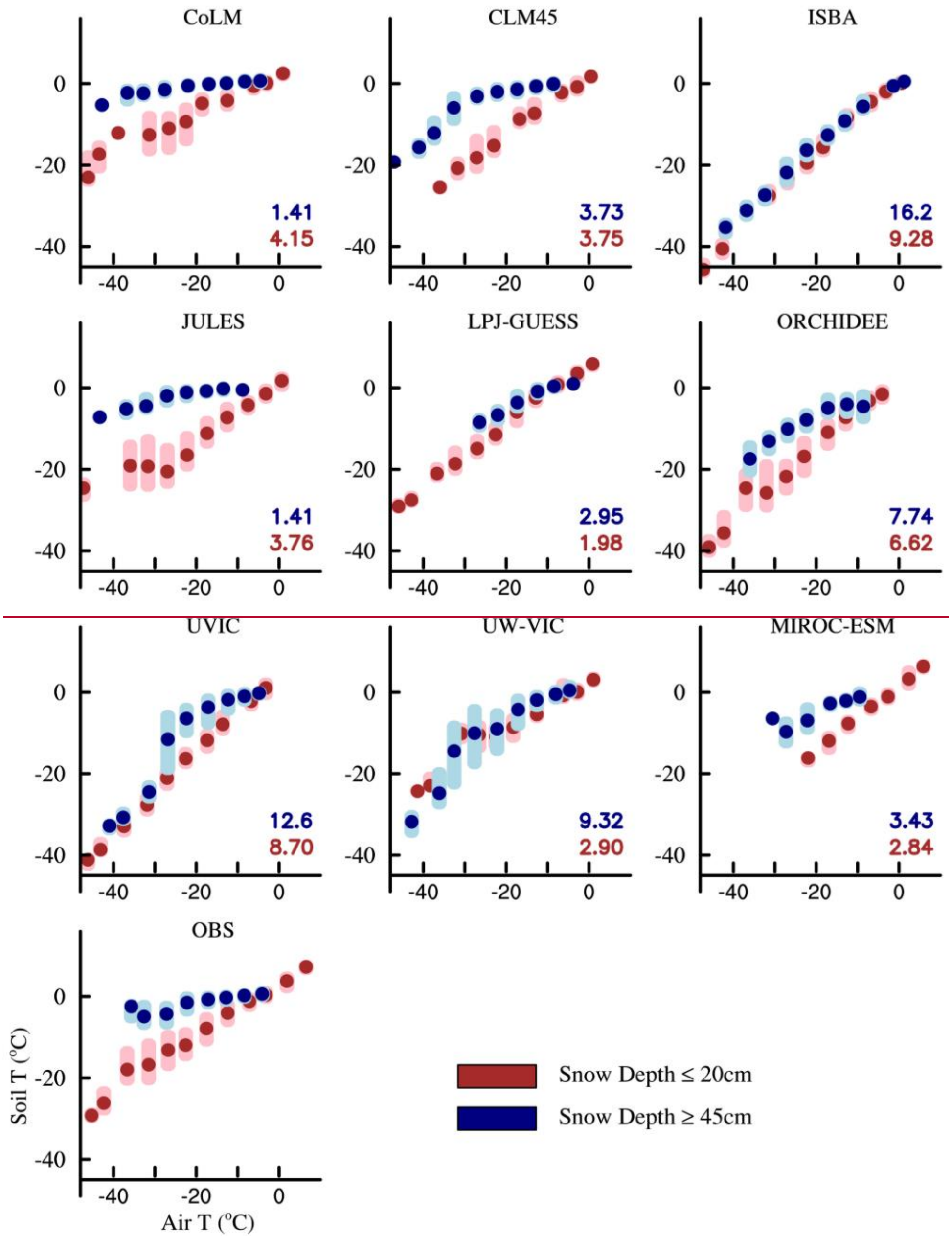


Figure 3. Spatial maps of the correlation coefficients between snow depth and ΔT, the difference between soil temperature at 20 cm depth and air temperature for winter 1980-2000. Regions with greater than 95% significance are hashed.

1 
CoLM
CLM45
ISBA
JULES
LPJ-GUESS
ORCHIDEE
UVIC
UW-VIC
MIROC-ESM
OBS
1.41
4.15
3.73
3.75
16.2
9.28
1.41
3.76
2.95
1.98
7.74
6.62
12.6
8.70
9.32
2.90
3.43
2.84
Soil T (°C)
Air T (°C)
Snow Depth ≤ 20cm
Snow Depth ≥ 45cm

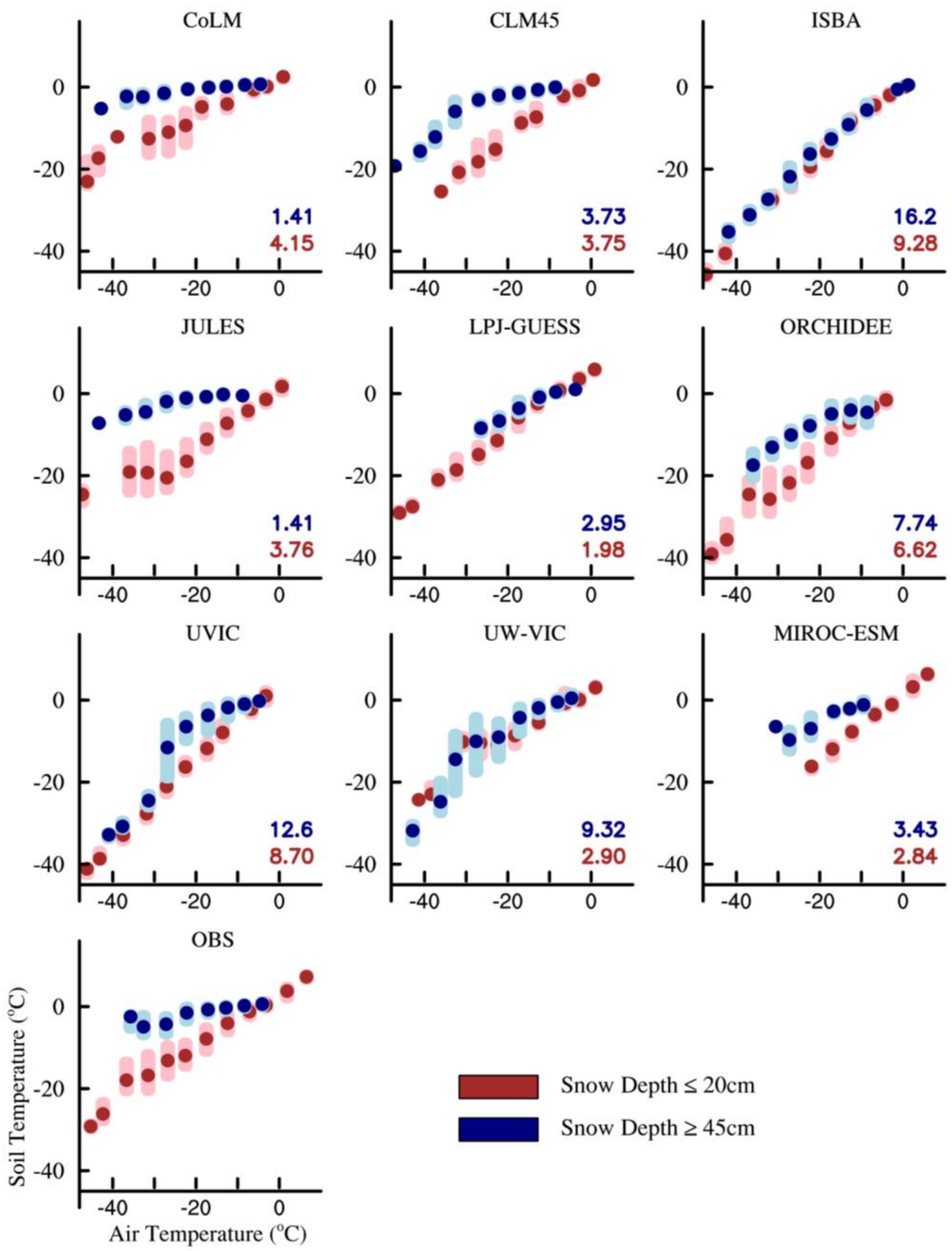
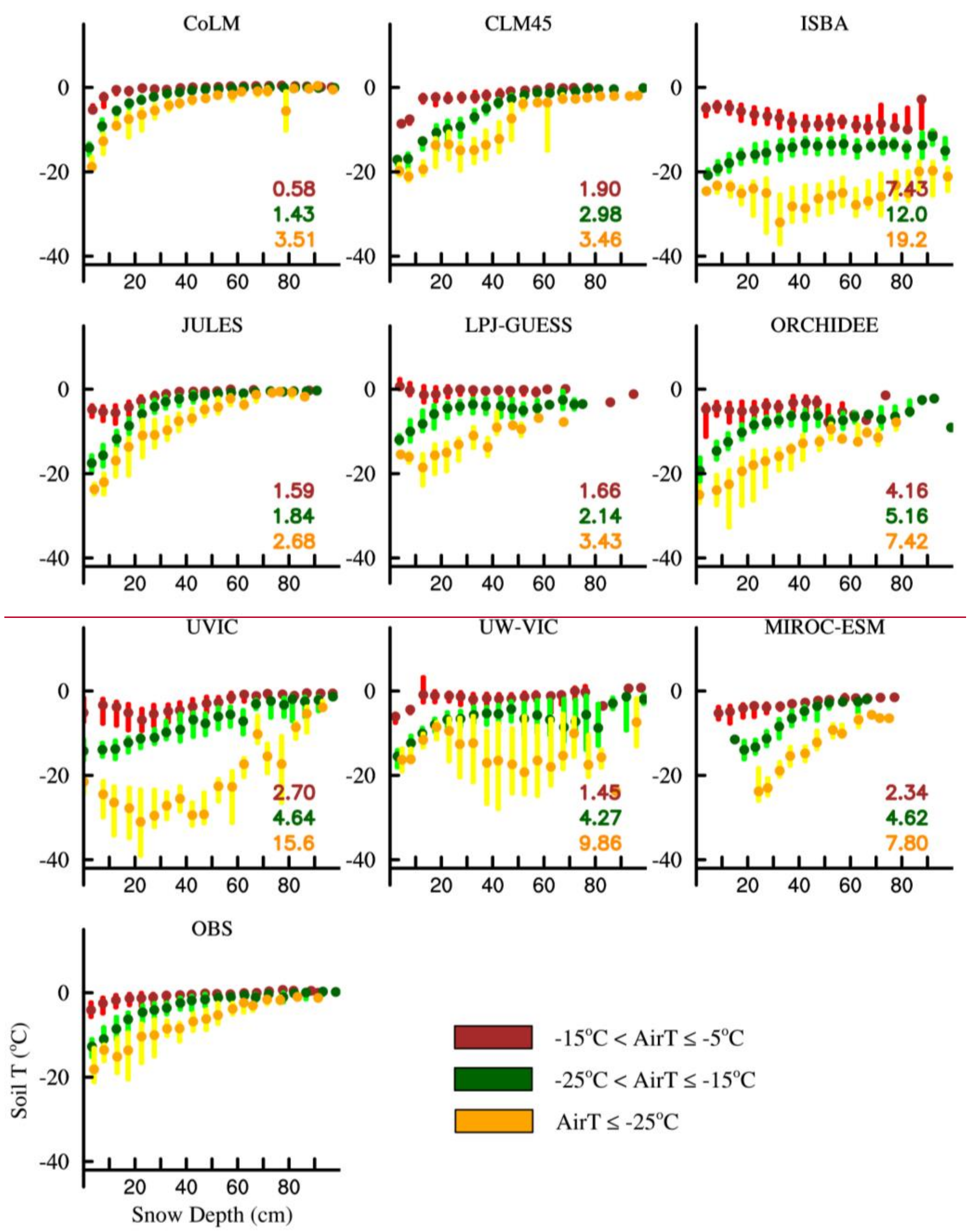


Figure 4. Variation of soil temperature at 20 cm depth (℃) with air temperature (℃) for winter 1980-2000. The dots represent the medians of 5 ℃ air temperature bins and the upper and lower bars indicate the 25th and 75th percentiles, calculated from all Russian station grid points (n=268) and 21 individual winters. The numbers in each model panel indicate the RMSE between the observed and modeled relationship. Color represents different snow depth regimes.

1 
CoLM
CLM45
ISBA
JULES
LPJ-GUESS
ORCHIDEE
UVIC
UW-VIC
MIROC-ESM
OBS
0.58
1.43
3.51
1.90
2.98
3.46
7.43
12.0
19.2
1.59
1.84
2.68
1.66
2.14
3.43
4.16
5.16
7.42
2.70
4.64
15.6
1.45
4.27
9.86
2.34
4.62
7.80
Soil T (°C)
Snow Depth (cm)
-15°C < AirT ≤ -5°C
-25°C < AirT ≤ -15°C
AirT ≤ -25°C

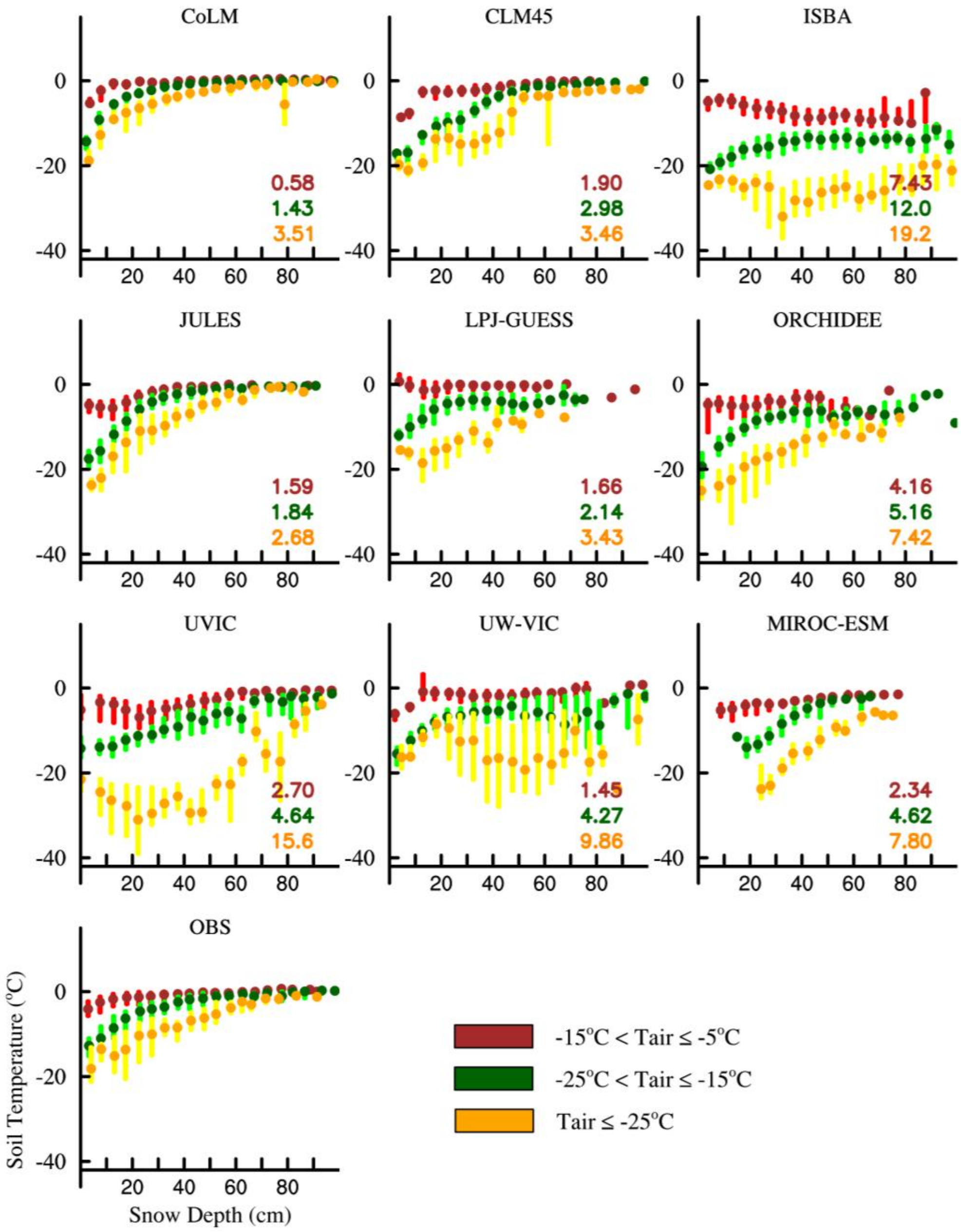


Figure 5. Variation of soil temperature at 20 cm depth (℃; y axis) with snow depth (cm) for winter 1980-2000. The dots represent the medians of 5 cm snow depth bins and the upper and lower bars indicate the 25th and 75th percentiles, calculated from all Russian station grid points (n=268) and 21 individual winters. The numbers in each model panel indicate the RMSE between the observed and modeled relationship. Color represents different air temperature regimes.

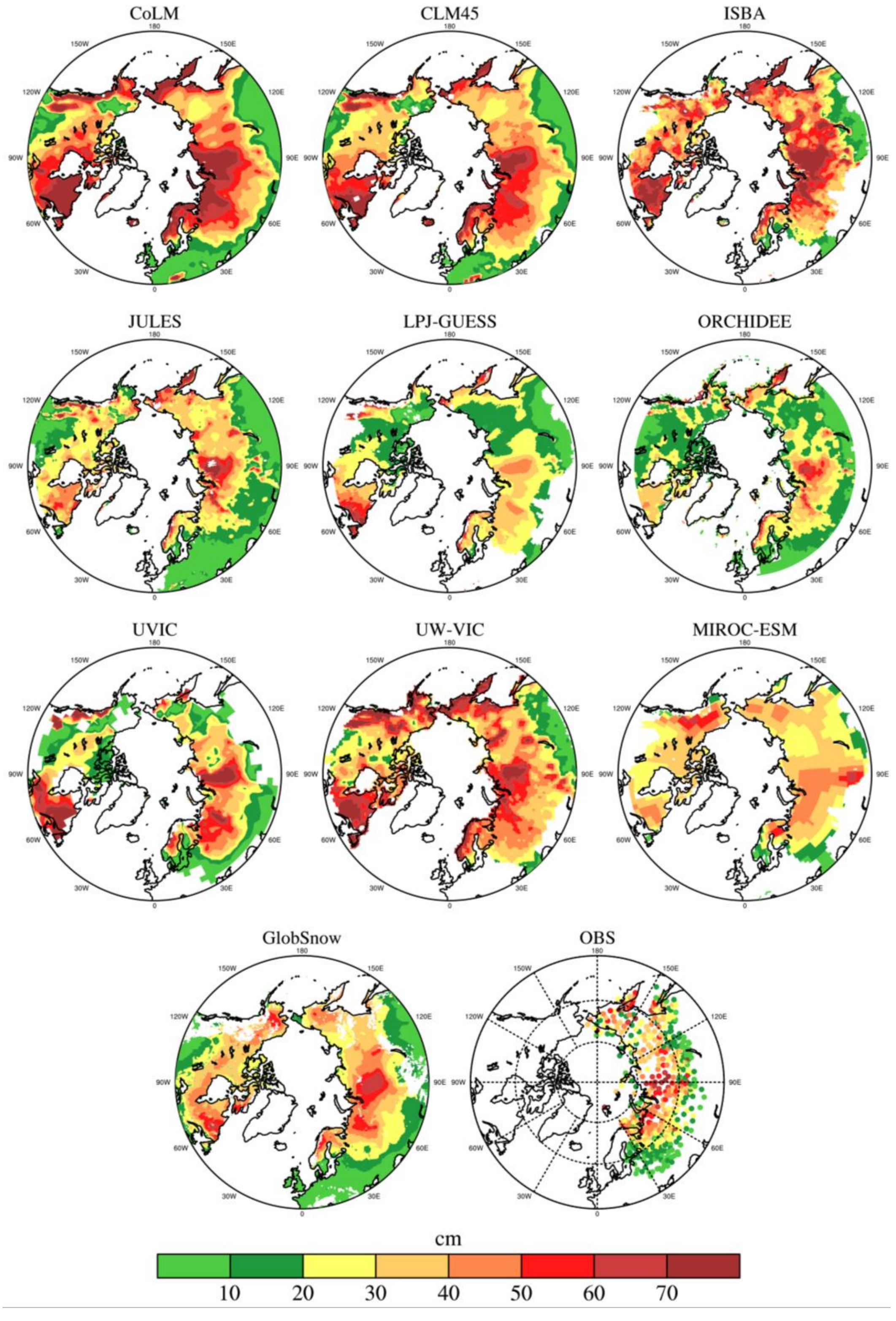


Figure 6. Spatial maps of snow depth (cm) for winter 1980-2000.