Response to comments of reviewer #1:

Major Comments:

1. This study used Wang06 permafrost map as a reference. As compared with Wang06map, there is a tendency that the indirect methods (MAAT, F, SFI) overestimated the permafrost area (Table 3). For the calculation, authors used thresholds of -2^{o}C MAAT and frost number ≥0.5. It seems their selection was considerably subjective, based on empirical results of previous studies. **The thresholds have considerable potentials that can change the results of this study.** For example, when MAAT is set to 0^{o}C, the permafrost area becomes probably larger than the areas indicated on Table 3. **That is, the results of this study also include latent uncertainties in terms of the methodologies.** A possible way to reduce the uncertainty is what examines the sensitivity of permafrost area against the changed threshold, calculating changes in the permafrost area to the ranges from -3 to 1^{o}C of MAAT. The calculations have to be made for frost number. The have to be summarized as a table and figure, including the discussions.

We have done as the reviewer suggested and added extra thresholds as a new column in Table 2 (i.e., -3 ℃<MAAT<0 ℃; 0.4<F<0.6; 0.4<SFI<0.6; 0 ℃<MAGT<0.5 ℃). And the derived permafrost area is now listed as extra rows in Table 3.

Generally, when the permafrost definition requires colder climate, the derived permafrost area becomes smaller. The across-threshold uncertainty (Table 3) is similar for different models. But the across-threshold uncertainty with SFI varies greatly among models, 23 ~ 105×10^{4} km^{2}, which is due to the seasonal amplitude of ground surface temperatures it requires. This is illustrated in Fig. 5 where UVic and LPJ-GUESS have a relatively small seasonal amplitude of ground surface temperature, which corresponds to their small across-threshold variability for SFI derived area in Table 3.

The across-model uncertainty is highly consistent even with different thresholds for each method (Table 3 final column). Thus it seems changing the thresholds does not affect one key point in our paper: that across-model uncertainties using direct methods are much larger than using indirect ones. Large across-model uncertainties using direct methods imply that differences among these land surface processes are worthy of investigation.

Accordingly, we add this analysis in the new discussion in section 6.1.

Revised Table 2.The five diagnostic methods and threshold values used to derive permafrost. The thresholds commonly used in the literature and in this paper are marked in bold.

Method	Definition	Threshold	Data used for calculation
TSL	More than 24 consecutive months soil temperature ≤ a threshold	**0 ℃**	0 ~ 3m monthly soil temperature
MAGT	Mean annual of 3 m soil temperature ≤ a threshold	**0 ℃**, 0.5 ℃	Mean annual of 3 m soil temperature
SFI	Surface frost number ≥ a threshold	0.4, **0.5**, 0.6	Annually maximum and minimum ground surface temperature
F	Air frost number ≥ a threshold	0.4, **0.5**, 0.6	Annually maximum and minimum air temperature
MAAT	Mean annual air temperature ≤ a threshold	0 ℃, -1 ℃, **-2 ℃**, -3 ℃	Mean annual of air temperature

Revised Table 3. Derived permafrost area inside the common modeling region on Tibetan plateau (10^4 km^2) from 6 LSMs and 5 diagnostic methods, using different thresholds. The results of thresholds commonly used in the literature and in this paper are marked in bold.

		CLM4.5	CoLM	JULES	UVic	ISBA	LPJ-GUESS	across-model uncertainty
Indirect method	MAAT⩽ 0 ℃	130	124	126	116	127	129	14
	MAAT⩽ -1 ℃	122	117	119	109	119	120	13
	MAAT⩽ -2 ℃	**113**	**105**	**111**	**99**	**109**	**110**	**14**
	MAAT⩽ -3 ℃	95	83	96	81	91	93	15
	across-threshold uncertainty	35	41	30	35	36	36	
	F⩾ 0.4	140	135	138	126	138	138	14
	F⩾ 0.5	**135**	**127**	**131**	**118**	**130**	**131**	**17**
	F⩾ 0.6	117	93	106	89	100	101	28
	across-threshold uncertainty	23	42	32	37	38	37	
Direct method	**TSL**	**60**	**1**	**62**	**8**	**44**	**119**	**118**
	MAGT⩽ 0.5 ℃	112	102	104	8	72	131	123
	MAGT⩽ 0 ℃	**104**	**89**	**96**	**8**	**61**	**128**	**120**
	across-threshold uncertainty	8	13	8	0	11	3	
	SFI⩾ 0.4	135	122	130	32	131	127	103
	SFI⩾ 0.5	**116**	**62**	**100**	**8**	**113**	**119**	**111**
	SFI⩾ 0.6	42	17	38	4	55	104	100
	across-threshold uncertainty	93	105	92	28	76	23	
across-direct method uncertainty (based on commonly used methods TSL, MAGT≤0°C, SFI≥0.5)		**56**	**88**	**38**	**0**	**69**	**9**	

2. a) Snow cover has the insulation effect on soil temperature; deeper snow depth increases soil temperature and vice versa. However, figure 7 and 8 have results inconsistent with the facts, especially found in ground surface temperatures of UVic, ISBA, and JULES. They are not common. The explanations authors mentioned are not enough to clear up the question.

We agree with the basic snow insulation mechanism, but UVic behaves differently from that of ISBA and JULES.

In the case of UVic, the ground surface temperature was warmer than the air despite of no snow cover.

There are two aspects that account for UVic behaviour: albedo and sublimation. Fig. 6 now more clearly shows that UVic simulates no snow (even in winter). Since there is no snow, the albedo is quite low all year round, 0.15-0.35. This means UVic can absorb much more solar radiation in winter, which can greatly warm the ground.

However, since the forcing of UVic is the same as CLM4.5, there is snowfall. Then where does the snow go? We assume the sublimation removes the snow in UVic. If so, it should take more energy to sublimate snow than it does to melt it, then the latent heat flux should be higher in UVic than other models. Actually we did find that the evaporation rate in UVic is higher.

In conclusion, the low albedo leads to more solar radiation absorption, which warms the ground and provides energy for sublimation. Although sublimation can cool the ground, the warmer ground indicates the low albedo effect is stronger in UVic.

We have addressed this in section 5.

In contrast, the surface temperatures of ISBA and JULES were colder under thicker snow cover.

Yes, observations show that in general deeper snow depth increases soil temperature. However, if the snow is too thin, the warming effect is very weak. But the melting, evaporation, and sublimation processes will take much energy from the ground, and the snow cover will reflect much solar radiation. Both will cool the ground. That is the thin snow cooling effect.

Fig.6 shows that in most places on Tibetan Plateau the snow depth of ISBA and JULES is less than 10 cm. Thus, the colder ground temperature of ISBA and JULES may be due to the thin snow. We investigate this in the new plots Fig. 9, which is like Fig.7 in the paper, but shows the temperature offset between ground surface and air temperature for different snow depths. By inspection we note that there is different behavior for snow depths thinner and thicker than 4 cm. For snow depth > 4 cm, most negative offsets disappear in ISBA and JULES, which means that the ground surface temperature is warmer than air temperature for snow depth larger than 4 cm. For snow depth < 4 cm, the ground surface temperature of much of the region is colder than air temperature in ISBA and JULES, which indicates the cooling effect of thin snow. The very small or slightly negative temperature offset for thin snow is also seen in the other models. Of course, the strength of this effect depends on the individual model's simulation/parameterization of the snow processes (such as sublimation, evaporation, melting). The thin snow mechanism is also confirmed by the weak insulation effect in Fig. 10.

Accordingly, we have improved the discussion of these issues in section 5. We also add Fig. 9 and Fig. 10.

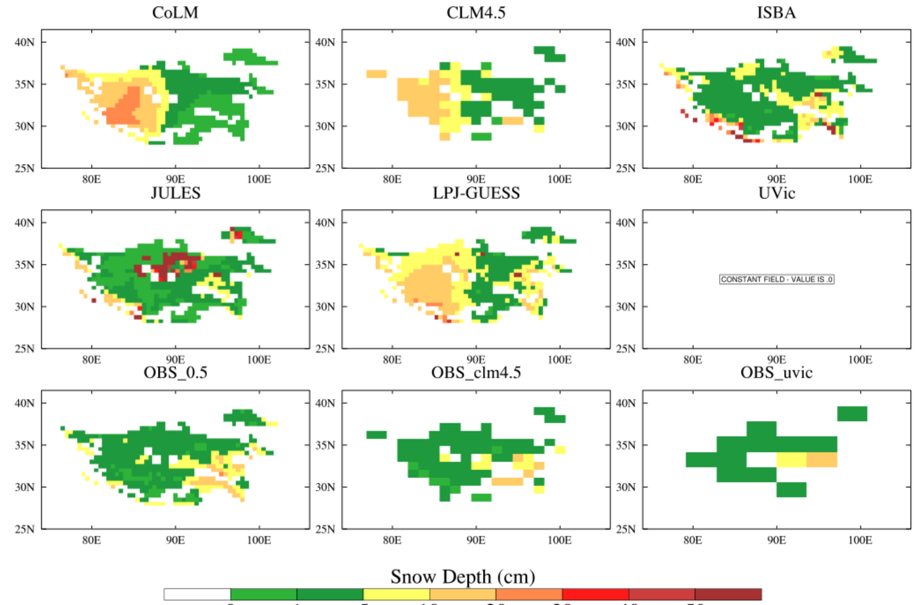


Figure 6. *Winter snow depth for the common region, averaged over 1980-2000. Note the nonlinear color scale. We use the Long Time Series Snow Dataset of China (Che et al., 2008) (http://westdc.westgis.ac.cn) as observed snow depth. The observed snow depth plot is further interpolated onto the models' resolutions as "OBS_". The OBS_05 is in 0.5 °resolution for CoLM, ISBA, JULES and LPJ-GUESS. The OBS_CLM4.5 and OBS_UVic are in the resolutions of CLM4.5 and UVic separately.*

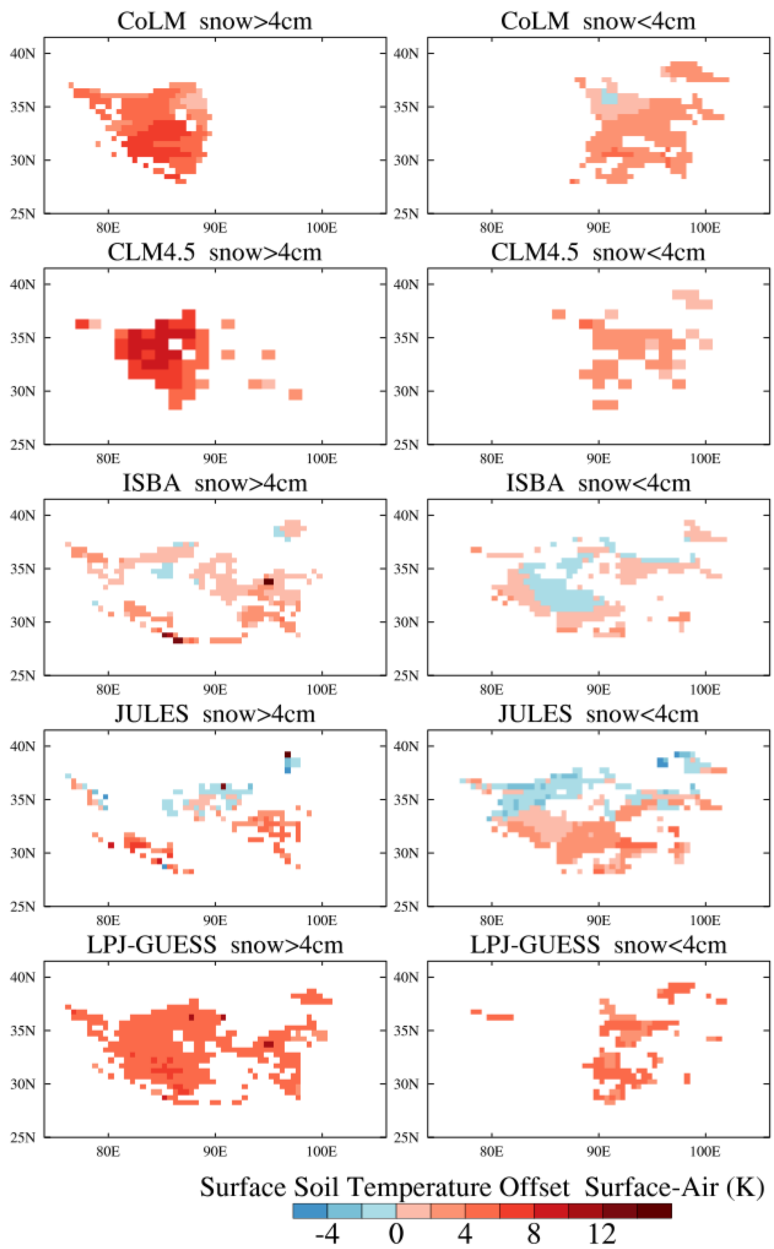


Figure 9. *Mean surface temperature offset (difference in mean winter temperatures between surface soil and air, averaged over 1980-2000). Left column is for snow depth >4cm, right column shows regions with snow depth<4 cm. Warm colors indicate soil is warmer than air temperature.*

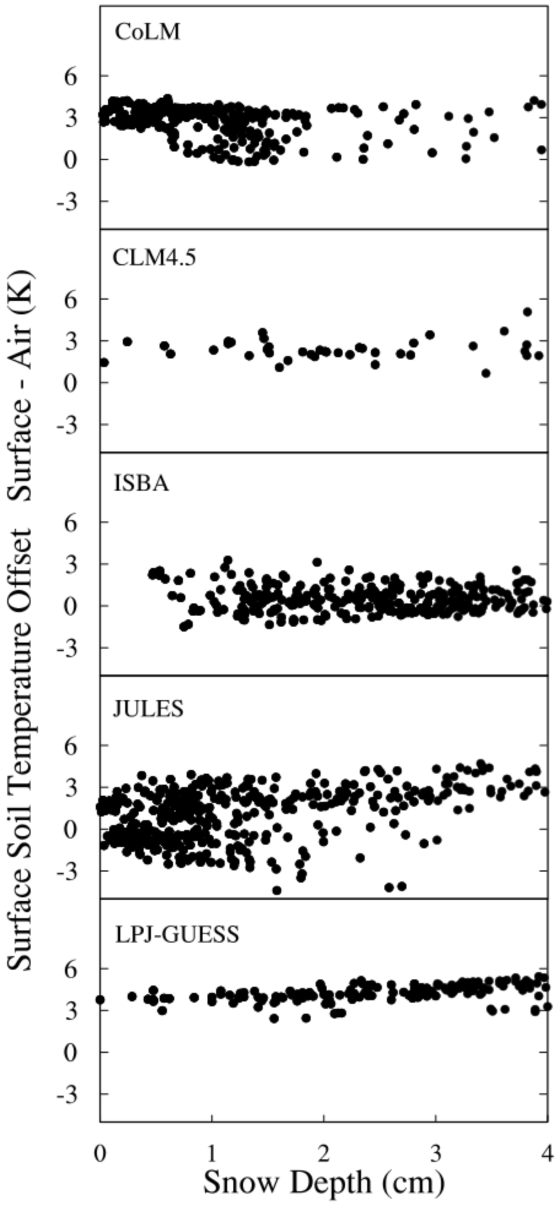


Figure 10. *Mean surface temperature offset (difference in mean winter temperatures between surface soil and air, averaged over 1980-2000) as a function of snow depth for grid points where average snow depth <4 cm.*

For the reasons, the authors mentioned problems of parameters the model used and the reliance of snow cover data derived from satellite observations. Of course, we can enough consider their influences. However, if there is problem in the snow cover data, how do you explain the results of the remaining three models that satisfy the common facts?

According to Wang et al. (2013), the snow depth pattern and the significant seasonal snow characteristics of satellite data are consistent with those of station data in most of our common TP region. The satellite data are different from station data on the southeast of TP (Wang et al., 2013), however, our analyzed common region does not include this part of TP. We add discussion in the second paragraph of Sect.5 where we introduce satellite data.

2. b) The ground surface temperature was extrapolated from the vertical soil temperature profiles. How did you extrapolate it, by liner or exponential way?

Ground temperatures were linearly interpolated onto the common depths: 0.05, 0.1, 0.2, 0.5, 1, 2, 3m. And we use the top two layers' soil temperature to linearly extrapolate the ground surface temperature. For CLM4.5, CoLM, ISBA and LPJ-GUESS, the first layer soil depth is no deeper than 0.01m and the second layer soil depth is no deeper than 0.05m. For JULES and UVic, the first layer soil depth is 0.05m and the second layer soil depth is no deeper than 0.18m. We do not expect that the presented results are sensitive to the interpolation procedure.

Accordingly, we improve these sentences in P1779 L10-2 as "... ...Ground temperatures were linearly interpolated onto the common depths... ...Since there is no ground surface temperature output, we linearly extrapolate the top two layers' soil temperatures onto the ground surface. For CLM4.5, CoLM, ISBA and LPJ-GUESS, the first layer soil depth is no deeper than 0.01m and the second layer soil depth is no deeper than 0.05m. For JULES and UVic, the first layer soil depth is 0.05m and the second layer soil depth is no deeper than 0.18m. Most TP permafrost work... ..."

2. c) Soil temperature doesn't generally have linear vertical profile. As the first step, you have to check the soundness for the extrapolated surface temperature; a way is to calculate the temperature offset using soil temperatures of 0.05 m. If there are no large differences, we could be doubtful about the snow cover data. If not so, you could have to recalculate the surface temperature that is implicated to SFI.

We agree. We plotted the surface temperature offset and soil temperature offset using the soil temperature of 0.05m depth, which can be compared with our Fig.7 and Fig.8, see below. The results show that first, the temperature offset patterns using 0.05m soil temperature are consistent with Fig.7 and Fig.8. Second, the offset between 0.05m soil and air temperature is only slightly different from Fig.7. Third, there is still negative surface offset in ISBA and JULES using 0.05m soil temperature as in Fig.7. This indicates that our interpolation for ground surface data is reasonable.

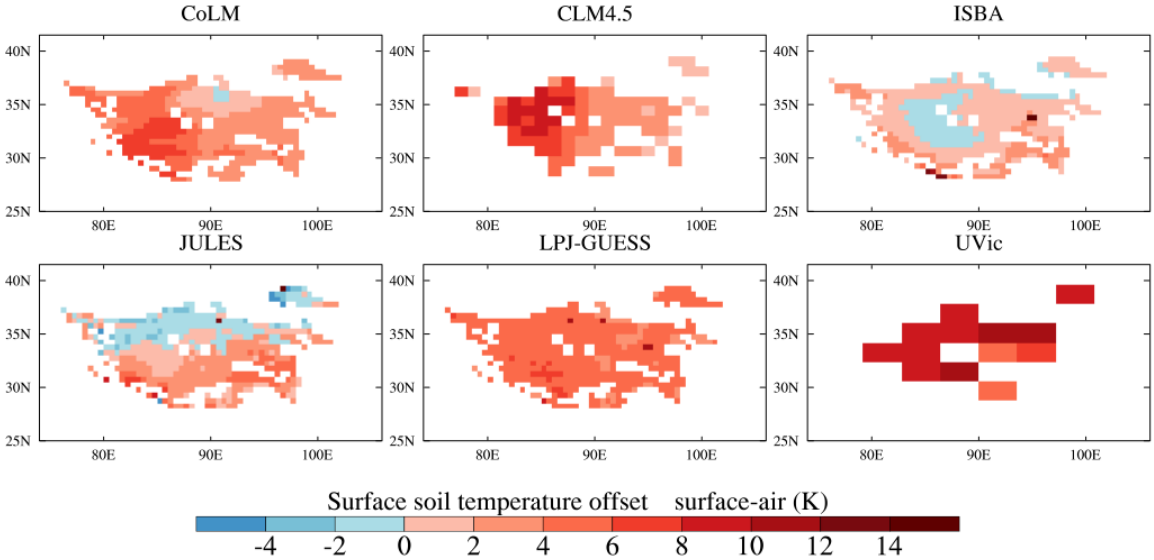
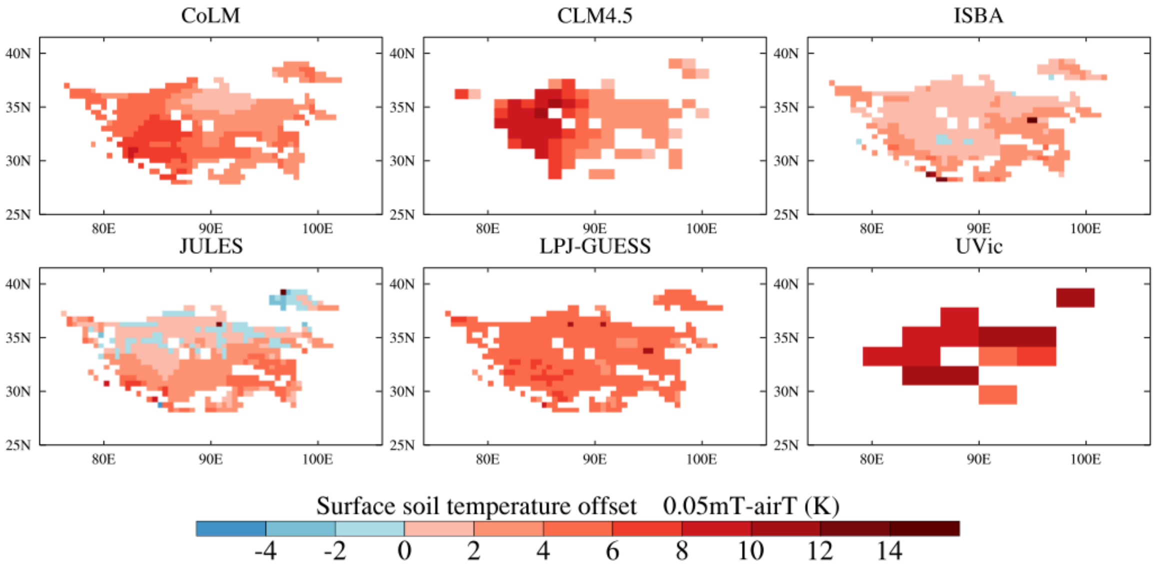


Figure 7. Mean surface temperature offset: difference in mean winter temperatures between surface soil and air, averaged over 1980-2000. Warm colors indicate soil is warmer than air temperature.



Mean surface temperature offset: difference in mean winter temperatures between ***0.05m soil and air****, averaged over 1980-2000. Warm colors indicate soil is warmer than air temperature.*

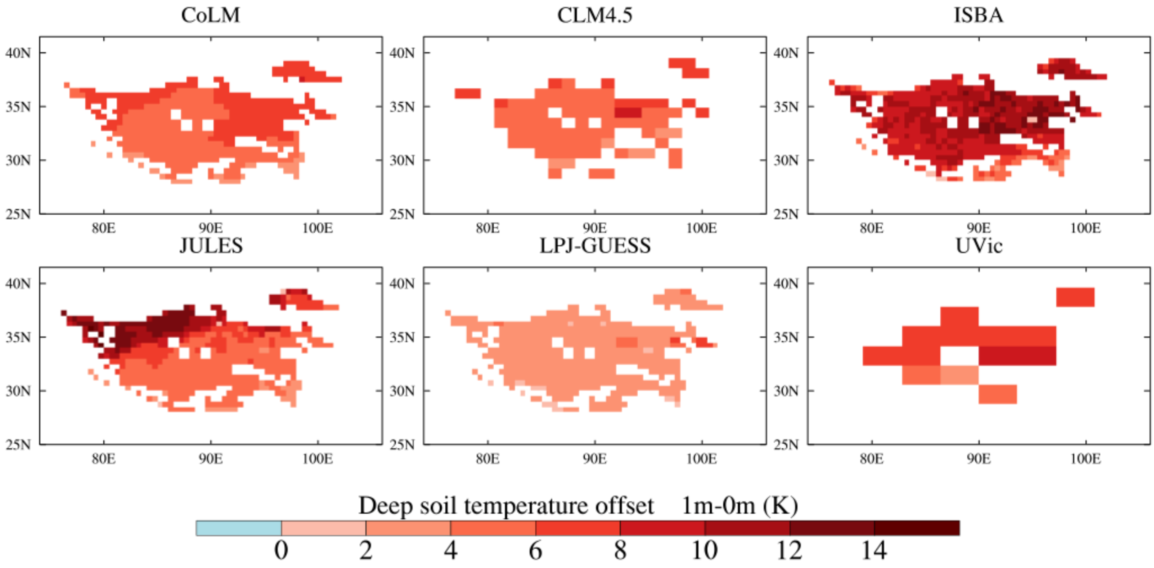
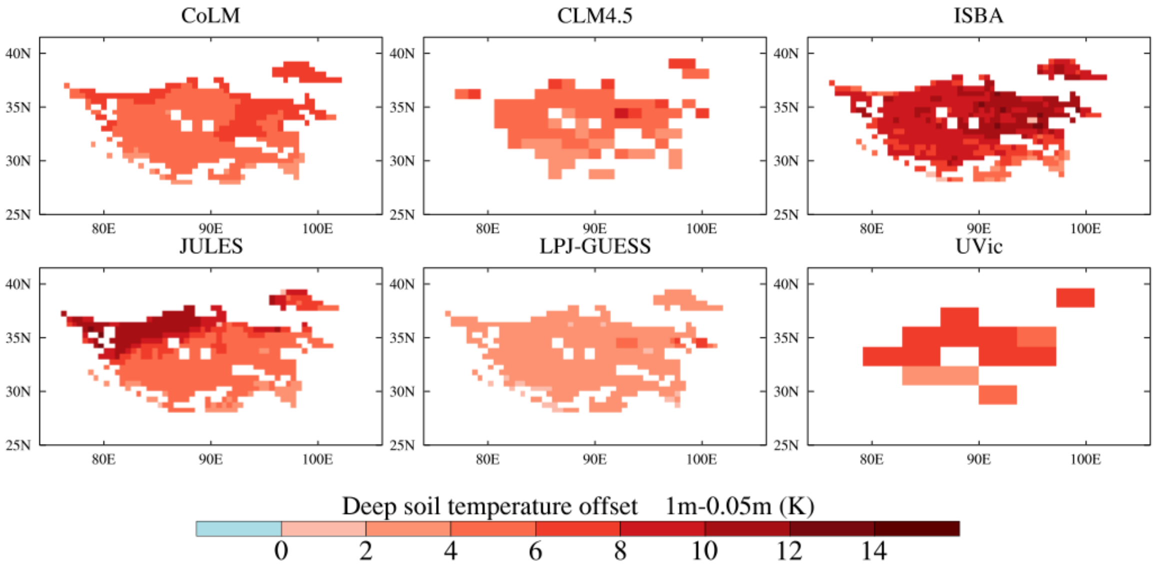


Figure 8.*Mean soil temperature offset: difference in mean winter temperatures between soil at 1 m depth and surface soil, averaged over 1980-2000. Warm colors indicate deep soil is warmer than shallow soil.*



Mean soil temperature offset: difference in mean winter temperatures between soil at 1 m depth and ***0.05m depth soil****, averaged over 1980-2000. Warm colors indicate deep soil is warmer than shallow soil.2.*

2. d) The authors mentioned influences of model parameters on the problem, but they didn't provide any scientific bases. For example, CLM4 (Lawrence et al., 2012, J. Clim.) and JULES (Dankers et al., 2011, Cryosphere) have characteristics of cold bias for soil temperatures. Likewise, you have to approach for the problem with scientific results and physical properties of individual models.

In this part we understand the referee means the issue of why UVic seemed to give large warming without snow, and also why ISBA and JULES give negative insulation for ground despite snow cover. We explained this in reply to part 2a above: it is due to albedo, sublimation and thin snow cooling effects. This is now explained in section 5, and it seems to us quite similar in method (that is using "scientific results and physical properties of individual models") as suggested by the referee and in the papers by Lawrence et al. and Dankers et al.

3. a) As the authors had already mentioned, the lowest soil boundary is a critical issue implicating to the permafrost uncertainty. Three models among the six models extended the soil to deeper depth, which makes it possible to do the discussion on this issue. The authors described a result of CLM4.5; when the soil was extended to 38 m, there was no significant change in the permafrost area (P1776 L3–5). From this description, the reviewer understands that there was a difference, though the difference was not so large numerically. If so, the authors have to provide the analyzed numbers on the manuscript, including results of other two models. The related description is summarized with one paragraph in the section of Conclusion, but which has to move to the Discussion section with additional discussion.

Agree, we move the paragraph to the discussion section and now provide the details of the calculations. As the reviewer suggested, we calculate the permafrost area with deeper soil layers. Note that it is the same simulation for the results of both 3 m and 30 m. "As UVic does not do a reasonable simulation of snow cover and ground temperature, we feel it is not necessary to include this model in the discussion here. Based on results from CLM4.5 and ISBA, the permafrost area calculated from MAGT at 3 m and at 10 m only changes by $1x10^4$ km^2. For results from CLM4.5, the areas calculated from MAGT at 20 m and 30 m do not change from the one calculated at 10 m. This is due to MAGT only considering annual mean soil temperature, not the seasonal cycle. This is consistent with the finding that the across-threshold uncertainty for MAGT-derived permafrost area is quite small (Table 3). However, the derived permafrost area with the TSL method improves when soil depth used for calculation is increased from 3 m to 5 m (Table 6). This sensitivity is because TSL requires information on the seasonal cycle of soil temperature. In other words, results of TSL method are sensitive to the active layer dynamics. The permafrost on TP is usually much warmer and has a deeper active layer than found in continuous permafrost of the arctic and boreal region. Hence deeper soil layers would be well suited for TP permafrost simulation. "

Accordingly, we add this analysis in the new discussion section 6.2, along with the new Table 6.

Table 6. Derived permafrost area (10^4 km^2)with deeper soil layers, using method "TSL". The results of thresholds commonly used in the literature and in this paper are marked in bold.

Depth of deepest layer used for calculation	CLM4.5	ISBA
3m	**60**	**44**
5m	85	54

3. b)The monthly time resolution this study used could also increase the uncertainties for the estimated permafrost area, which has also to be mentioned in the Discussion section.

Agree. The methods used here are defined for monthly data, but standard definitions are for daily data or hourly data. We do not have that output from the models. We added a sentence in the last of section 6.2: "Daily and hourly temperature data may make some differences to the permafrost extent map since the depth of diurnal cycle wave in soil is much less than the deepest model layer."

<u>Minor Comments:</u>

P1771 L9, the word 'best' is not appropriate. Please remove it.

Done. **It reads now** "There is good agreement (99–135 $\times 10^4$ km^2) between the two diagnostic methods based on air temperature which are also consistent with the observation-based estimate of actual permafrost area (101 $\times 10^4$ km^2)."

P1771 L15-7, it seems differences in vegetation types are implicated to.

Done. We add vegetation types. **It reads now** "Model evaluation at field sites highlights specific problems in process simulations likely related to soil texture specification, vegetation types and snow cover."

P1772 L4, climate likely affects the permafrost distribution.

Done. We add climate factor. **It reads now** "The unique geography and plateau climate make the permafrost on TP very different from the Arctic."

L10, does 'the temperature' mean air temperature?

Here we refer to the underlying surface temperature contrast. **It reads now** “The underlying surface temperature contrast between TP and Indian Ocean is an important controlling factor for both the Asian monsoon, and the wider general atmospheric circulation (Xin et al., 2012).”

L19, where are individual station locations?

The station of VIC model here is on the permafrost of Heilongjiang. Others are all on the Tibetan Plateau. We delete VIC model here. **It reads now** “A number of land surface models (LSMs) (e.g., CLM4.0, CoLM, SHAW, Couple Model and FSM) have been applied at individual station locations on the Tibetan Plateau to reproduce soil thermo-hydro dynamics (Li et al., 2009; Wang and Shi, 2007; Xiong et al., 2014; Zhang et al., 2012).”

P1774 L18-9, please add references.

Done. The references have been **added as** “We make use of all five major permafrost diagnostic methods promoted in the literature (Slater and Lawrence, 2013; Guo et al., 2012; Guo and Wang, 2013;Wang et al., 2006; Wang, 2010; Nan et al., 2002;Nan et al., 2012; Saito, 2013; Ran et al., 2012; Wang et al., 2006; Jin et al., 2007;Xu et al., 2001; Nelson and Outcalt, 1987).”

P1778 L15, what means ‘SD’?

Sorry, ‘SD’ means standard deviation. We **correct it as** “rather than defining uncertainty in terms of standard deviation”.

P1778 L26–P1779 L6, I don’t well understand why Miroc-ESM was used in this study.

This is a misunderstanding. We analyzed Miroc-ESM since Miroc-ESM is also under “the Permafrost Carbon Research Network”. But we didn’t put its results in the paper, since Miroc-ESM is totally different from the land models. See P1779 L1 “We also analyzed (but do not show here) out put from a coupled earth-system model (Miroc-ESM)”.

Since these sentences are rather misleading, we **deleted them** (P1778 L26 - P1779 L6) “We also analyzed (but don’t show here) ... ... with the stand-alone LSMs”.

Figure 5 certainly includes data of Miroc-ESM, but the data are not closely related to this study. Your intent is to strengthen the high uncertainty in the permafrost processesbetween models, is it?

Sorry, this is a mistake. Yes, the intent is to strengthen the high uncertainty in the permafrost processes between models. We **deleted** Miroc-ESM in Figure 5.

P177**8** L10-2, you have to describe more specifically the way on how the ground surface temperature was extrapolated.

We think here you mean we should improve the extrapolation method description in P177**9** L10-2. Please see our detailed reply to your question 2 b).

P1779 L17-24, it may be helpful if you describe the specific numbers about TP permafrost area reported by previous studies.

Agree. These numbers can further indicate the permafrost area uncertainty of the observation-based maps. We **add a sentence** in P1779 L24 **as** "...... different studies (Ran et al., 2012). Thus there is a large spread of observation-based TP permafrost area estimates from 110×10^4 km^2 (Wang et al., 2006) to 150×10^4 km^2 (Shi and Mi, 1988; Li and Cheng, 1996)."

P1781 L20-1, The description about Figure 5 is too simple. You have to add more explanation about data displayed on the figure. From the figure, we can see differences in the simulated soil temperatures between models. The air temperature also shows differences between the models, especially in winter season, though the differences are smaller.

Done. We add a sentence "The air temperature also shows differences between the models, especially in winter season, though the differences are much smaller than differences in soil temperatures."

P1782 L10, does 'observation' mean Wang06 map?

Sorry, this is a mistake. We change "observation" to "Wang06 map".

L19-21, we can consider the impact of different forcing data, especially air temperature as identified in Figure 5. Therefore, it can't conclude the difference with just different spatial resolution.

Agree, and delete the sentence.

P1786 L20-1, for 'poor representation of soil hydrology', I can't find any scientific results from this paper to support the description.

Well, here we think the soil hydrology of LPJ-GUESS plays two parts in underestimating of soil temperature 1) through affecting soil thermal properties, see P1786 L24-25 "This suggests a different (larger) winter soil thermal conductivity probably associated with a high soil porosity and water content"; 2) through hydrological process, e.g. ice melting in summer, which can be confirmed by Fig. 4a and c, see P1787 L2-7 "Precipitation and hydrological processes determine the vertical profile of soil water content which can change the fraction of water and ice retained in different soil layers and influence soil thermal conduction. The energy required to melt the high water (ice) content in the surface soil layers in summer appears to lead underestimated low summer temperatures compared with other models, and a phase lag in summer warming (Fig. 4a and c)." Thus we think "poor representation of soil hydrology **on Tibetan Plateau**” is solid with the two parts of soil hydrology in LPJ-GUESS.

P1788 L14-6, is the description able to apply to all models?

In fact, CLM4.5, CoLM, JULES and ISBA do produce better permafrost maps with MAGT and SFI than TSL. LPJ-GUESS and UVic are two exceptions, showing us no change from TSL to MAGT and SFI. Here the word “generally” means “most models”. But if this is misleading, we would change it.

Accordingly, **we change this sentence with** “Most models in this study produced permafrost maps in better agreement with the Wang06 map using the MAGT and SFI methods rather than with the TSL method.”

Table 3, it may be helpful if the uncertainties compared to Wang06 are included.

We find that showing the uncertainties compared to Wang 06 in Fig.2 works better than in this table.

Figure 5, it may be helpful if the seasonal snow depth derived from the data of Figure6 is added.

We add additional discussion and extra figures on snow depth that we think better explains the points. We address this in detail in our reply to your second comment.

Response to comments of reviewer #2:

Major Comments:

1. I agree with the other reviewer, that the snow issue should be better investigated. The results shown here are against many other publications stating the effect of snow insulation on soil temperature.

We refer the referee to our answer to question 2 of referee #1 where we show that the results are not in fact against previous publications, but fully in line with them when the effects of albedo and thin snow covers are considered.

2. One suggestion I can make is to calculate a "snow season" instead of using DJF values since snow can be persistent over spring. You can simply do it by air temperatures [days Tair <0] or if available, model show depth data [days snow depth >a threshold value like 1-5 cm]. Then compare the air vs surface temperature offsets during this snow season and see the results.

Unfortunately daily modeling output data under the Permafrost Carbon Research Network is not stored.

However we agree it is a good idea to use snow depth thresholds, using 4cm snow depth to explain the mechanisms of ISBA and JULES. See figures 9 and 10 of the revised paper and the detailed reply to question 2 of referee #1.

3. There needs to be subsection describing each model used in this study. It doesn't has to be long but at least give some important details about which processes they utilize and what major differences (grid-size/soil discretization/physical-biogeochemical processes/ snow schemes etc.) they posses compared to other models here.

We add the following paragraph in section 3.1.

"The LSMs in this study considered the following processes: dynamic vegetation, carbon cycling (Rawlins et al., 2015), snow, near-surface hydrological budget, soil thermal dynamics (Peng et al., 2015) and the treatment of freezing soil. Sophistication in the treatment of these processes varies amongst the models with each having specific parameterizations, In this study we investigate some key schemes and parameters that are important for permafrost simulation: 1) Unfrozen water / phase change. All models calculate soil thermal properties as a function of soil moisture and consider the phase change of water/ice, but CoLM and LPJ-GUESS do not consider transformation to ice of water solute mixtures below 0 ℃, which is a key feature in soil freezing and thawing. 2) Surface organic layer insulation. Only CLM4.5 and ISBA consider the insulating effect of moss. 3) Soil texture parameterization. The specified fraction of clay and sand in soil differs. LPJ-GUESS specifies the same soil texture for the TP as for the Arctic. 4) Organic soil fraction treatment. The organic content of soil differs among the models. LPJ-GUESS sets the same value for TP as for the more organically rich permafrost of the Arctic. 5) Snow processes. ISBA, LPJ-GUESS and UVic set static snow layers. UVic uses an implicit snow scheme while LPJ-GUESS uses the Bulk-layer scheme, which are both simpler than the dynamic multi-layer snow scheme of some other land models."

We also improve the Table 5 accordingly.

Revised Table 5. Year-round relative model characteristics on TP.

Model	Snow cover[1]	Albedo[2]	Soil water[3]	Unfrozen water effect during phase change[4]	Surface organic layer insulation	Snow scheme[5]
CLM4.5	Medium	Medium	Medium	Yes	Yes	Dynamic & ML
CoLM	Medium	Medium	Medium	No	No	Dynamic & ML
ISBA	Low	Low	Medium	Yes	Yes	Static & ML
JULES	Low	Low	Medium	Yes	No	Dynamic & ML
LPJ-GUESS	Medium	Low	High	No	No	Static & BL
UVic	None	Low	High	No	No	Static & L

[1] Low snow cover is confined to high elevations, medium tends to be on western TP

[2] LPJ-GUESS has constant albedo everywhere and UVic albedo varies slightly due to vegetation, year-round albedo variability for other models depends mainly on snow cover in winter and soil moisture, vegetation, etc in summer

[3] soil water content includes both liquid and ice fractions

[4] all models calculate soil thermal properties depending on soil moisture and also phase change of water, but CoLM and LPJ-GUESS ignore solute dependent freezing processes

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

Also some reference papers for each model should be included.

We add references for models in Table 1.

Table 1. The six land surface models, analyzed over the Tibetan plateau (TP).

Model	Native Resolution	Number of soil layers	Depth of soil column (m)	Spatial domain	Atmospheric Forcing Data
CLM4.5 Swenson and Lawrence, 2012 Oleson et al., 2013	1°×1.25°	30	38.1	Whole TP	CRUNCEP4[1]
CoLM Dai et al., 2003	1°×1°	10	2.86	Whole TP	Princeton[2]
ISBA Decharme et al. 2011	0.5°×0.5°	14	10	Permafrost region follow IPA map	WATCH [3]
JULES Best et al., 2011	0.5°×0.5°	30	2.95	Whole TP	WATCH [3]
LPJ-GUESS Gerten et al., 2004 Wania et al., 2009	0.5°×0.5°	25	3	Permafrost region follow IPA map	CRU TS 3.1[4]
UVic Meissner et al., 2003	1.8°×3.6°	14	198.1	Whole TP	CRUNCEP4[1]

[1]Viovy and Ciais (http://dods.extra.cea.fr/)
[2]Sheffield et al. (2006) (http://hydrology.princeton.edu/data.pgf.php)
[3]Weedon et al. (2011) (http://www.waterandclimatechange.eu/about/watch-forcing-data-20th-century)
[4]Harris et al. (2013), University of East Anglia Climate Research Unit (2013)

4. Same for the 3 site locations. There is already good information on Table 4 about the sites but still it will be good to include a small subsection describing the similarities/differences among these sites. Especially at Fig.4, the 0.04m observation of D105 and 2.63m observation of D110 are missing. You can better explain the reasons in a subsection. Especially at Fig.4, the 0.04m observation of D105 and 2.63m observation of D110 are missing. You can better explain the reasons in a subsection.

We add sentences in P1781 L15, "... model results. The three stations are located along the Qinghai-Tibet Highway. D66 station is in the front edge of alluvial fan, with almost no vegetation. The soil is mainly composed of gravels, sands and pebbles. D110 is in the southern bank of ZhaJiaZangBu River. The ground is a wetland covered with short-stature emergent vegetation. The upper layer soil is composed of coarse and fine sand. The lower soil layer is mainly composed of fine sand. D105 is in the northern side of the Tanggula Mountain range. The ground surface is relatively flat, covered with plateau meadow. The soil is composed of both coarse and fine sand. The vertical profile of observed soil temperature of D66 extends from 0.04 m to 2.63 m, of D110 from 0.04 m to 1.8 m, and of D105 from 0 to 3 m. However the data continuity of the top layer temperature in D105 is not good. To examine modeled ground temperatures, we present the top ... ..."

and why you choose to compare these sites.

We have only the 3 observation sites and we have addressed this in P1781 L12-L13 "..., field studies on TP are quite limited, and we have only short duration (1996-2000) ground temperature profiles obtained from the GEWEX Asian Monsoon Experiment..."

Also, you can explain that you have used cutout of global simulations instead of running the models with the observed forcing for these sites and its consequent implications to the results.

We have addressed this in the 3.4 section, see P1781 L16-L17 "...temperatures (modeled temperatures were weighted bilinear interpolated onto the station locations) in Fig.4 and Table 4..."

5. I understand that the model results are gathered from RCN database and are restricted to the procedure of that project. However, monthly soil temperatures are not always enough for TSL style permafrost calculations. You can either request daily results from the modeling groups or at least mention this fact as one important reason for the performance of TSL method.

Unfortunately daily modeling output data under the Permafrost Carbon Research Network is not stored. Actually we have mentioned this point when addressing TSL method in the text. See P1775 L13-L18.

6. Your calculations are limited by model soil depth (3m) and you have mentioned that shortly in your text. However you can make more analysis with the models that have deeper soil layers. And maybe transfer the soil depth paragraph from conclusions to discussions. I leave this issue to the authors' choice.

This question is essentially the same as referee #1 question 3 to which we refer for our detailed answer. We address the point in the paper with a new Table 6 and discussion in a new section 6.2.

7. Observational map has its own uncertainty originating from the MAGT and statistical extrapolations. This should be mentioned more precisely in the text especially in your discussions.

Actually there is a paragraph in section 3.2 to specially address the uncertainty of Wang06 map. See P1780 L14-L18.

It is also good to address the uncertainty among different permafrost maps. We do this by describing the specific numbers about TP permafrost area reported by previous studies. We add a sentence in P1779 L24 as "...... different studies (Ran et al., 2012). Thus there is a large spread of observation-based TP permafrost area estimates from 110×10^4 km^2 (Wang et al., 2006) to 150×10^4 km^2 (Shi and Mi, 1988; Li and Cheng, 1996)."

To lower the impact of mismatches to Wang06 map, you might consider discussing inter-model range of TP permafrost area more. Fig.2b, for example, gives too much impact on mismatching Wang06 map.

The Wang06 map was re-gridded onto each model's resolution. To compare between the models would require each model to be re-gridded to each other's model resolution before comparison can be made. This would be confusing to describe. We also think that there is a fundamental difference in comparing models with an observation-based map rather than simply between each other.

8. To improve the scientific value of your model inter-comparison results you have to tackle each of the following issues: 1.forcing data, 2.model spatial resolution, 3.model time step, 4.model spin-up, 5.model soil layer dicretization,6.model soil depth, and finally 7.model processes. I assume it is most valuable to confine the differences to model processes and for that, one needs to make sure the others are the same or at least they have negligible differences. From your experiment, I see that only point 3 (time step) is the same. And you have mentioned point 1 and point 6 in your text. Although you have shown points 2 (spatial resolution) and 5 (soil layer discretization) in your Table 1, you did not mention them in the discussions.

So you should clear the issues regarding to points 2,4, and 5.

For point 2, we reduce the impact of spatial resolution on the results by re-gridding Wang06 map onto each model's spatial resolution. Although the spatial resolution varies among these models, the models are evaluated objectively and separately by comparing with Wang06 map of the their own spatial resolution, when comparing permafrost area and calculating kappa coefficients.

For point 5, we interpolate and extrapolate the ground temperature onto the common layers: 0, 0.05, 0.1, 0.2, 0.5, 1, 2, 3m. So there is no difference among models in calculating permafrost area when referring to soil layer discretization.

For point 4, model spin-ups could be of course different, but in all models were run for long enough (around 1 000 years) to ensure that the deep carbon is in equilibrium. Accordingly, we add one sentence in P1778 L25 to address this point, as "Model spin-ups are also different, but as they are long enough to ensure that the deep carbon is in equilibrium, about 1000 years, the spin up impacts should be small...."

9. Would it be possible (or useful) to include the correlation coefficients next to kappa metric?

We think kappa coefficient is already a good and common applied method for mapping similarity. So it is not necessary to use other statistics here.

I can suggest you to prepare a soil temperature plot showing annual mean, minimum, and maximum values of each soil layer temperature at the sites and maybe also the selected region or common region. With the soil temperature envelopes plotted in this style, we can see the mismatches of each model more clearly than the time series plots in Fig.4 and Fig.5.

Our plots of annual cycle in Fig. 4 and 5 show not only the maximum and minimum yearly temperatures (as requested) but also usefully show the time delays of models. And we compare the annual mean, and seasonal cycle amplitude of soil temperatures in detail in Table 4.

We improve the beauty of Fig. 4.

10. Why is LPJ-GUESS always the coldest? Your explanation in Sect 5 is very hypothetical. Unless you have the actual soil conductivity values or soil water content to compare, these are just candidates for the mismatch. This might as well be related to other soil processes like **type of soil heat transfer, coupling of soil water and vegetation cover** and several other soil parameters.

We agree other mechanisms you list are possible, and mention them as possibilities in the text. Unfortunately, the soil moisture output of LPJ-GUESS includes both liquid water and solid ice, not separately, and there is no soil conductivity output.

That must be one simple process that is uniquely different than other models.

As we have discussed, the underestimation of soil temperature in LPJ-GUESS can be due to many factors, and the most possible are inappropriate prescriptions of soil thermal properties, poor representation of soil hydrology, and mis-match of vegetation types **for the Tibetan Plateau**. See P1786 L18 - P1787 L7. Fig. 8, Fig. 4a and c, and the soil texture setting of LPJ-GUESS (Table 5) which all support our explanation.

One obvious problem for this model's results is that why is it colder even though it has higher snow depth.

We think that the cold ground temperature of LPJ-GUESS can also be explained by its snow depth, snow density and snow scheme. Accordingly, we added one paragraph to address this point in section 5. "LPJ-GUESS shows a similarly thick snow depth in the western part of Tibetan Plateau as CLM4.5 and CoLM (Fig. 6), but does not show as large surface temperature offset as those two models (Fig. 7). That is because LPJ-GUESS has a fixed snow density (362 kg/m3) which is higher than used in other models, and a relatively simple Bulk-layer snow scheme, with one static snow layer, unlike the dynamic multi-layer snow scheme of CLM4.5 and CoLM (Table 5)."

11. UVic is the warmest among models. You say UVic has no snow cover, then what is shown in Fig.6?

Please notice the color bar in Fig. 6 is not linear. The color in UVic represents some values smaller than 0.1. Actually the snow depth of UVic is less than 0.001 cm. But UVic does have snowfall and a simple snow scheme. Thus we treat the snow depth of UVic as 0. We improve Fig.6 to avoid such misinterpretation.

This is one other reason to explain models in a different section. You attribute the overestimated soil temperatures of UVic to snow sublimation. Then I don't understand why the soil is warmer. The longwave radiation should be used for this sublimation you mention, not to warm the soil. And since there is less snow cover in UVic (Fig.6), we should expect cooler ground temperatures, which is not visible in Fig 7.

Yes. The sublimation will cool the ground. But since there is no snow in UVic, the rather low albedo in winter (0.25) can absorb much more solar radiation, which will provide energy for sublimation and warm the ground. Although sublimation can cool the ground, the warmer ground indicate the low albedo effect is stronger in UVic.

Please see our detailed reply to question 2 of Referee #1.

12. I also don't understand the explanation of JULES and ISBA models being cooler at the surface even though they have much deeper snow depths.

Fig.6 shows that in most places the snow depth of ISBA and JULES is less than 10 cm. We argue this cooling is due to the "thin snow cooling effect".

Please see our detailed reply to question 2 of Referee #1 and figures 9 and 10.

13. What is the point of using MIROC-ESM results in this intercomparison? I don't see an immediate relevance comparing a fully coupled model to offline simulations of different models. Please justify your choice or remove that model.

We delete Miroc-ESM in Fig.5. And to avoid misinterpretation, we deleted the sentences (P1778 L26 - P1779 L6) "We also analyzed (but do not show here) ... ... with the stand-alone LSMs".

14. What is the message to model developers for a better TP estimate? What needs to be improved according to your results?

We address this in the conclusions: "Although most models can capture the threshold value of MAGT and SFI, their ground temperatures still show various biases, both in the mean annual value and the seasonal variation. Therefore, most models produce worse permafrost maps with the TSL method. The TSL method is a more demanding, and to date, elusive target" and "If the observation sites for soil temperature are representative, then LPJ-GUESS and UVic have substantial biases in their soil temperature simulations, mainly attributable to inappropriate description of the surface (vegetation, snow cover) and soil properties (soil texture, hydrology). Other models (ISBA, JULES) show biases in the simulation of winter soil temperature"

Minor Comments:

P1771 L17 and L20: produce "better" permafrost maps of the TP you mean?

Here we mean with MAGT and SFI 1) the derived permafrost area is nearer to observation-based estimate and 2) the permafrost distribution is also better, with higher kappa coefficient. We have addressed this through Sect.4.

P1772 L9: lose the comma

Done.

L9: "plays" -> "play important roles"

Done.

L11: lose the comma

Done.

P1773 L20: majority of your models must be tuned for several different sites around the world. What do you mean "different from where they were tuned"? Maybe you can mention that they are mostly used to estimate Arctic permafrost and not the TP. That can clarify the aims of this work. But these are global models and they are not tuned only to NH areas...

Agree. We rephrase the sentence as "We note that this approach provides information on the modeling ability of current models on the warmer and physically unique TP permafrost in a NH simulation, hence providing a test of reliability for simulations of present and future global permafrost over TP."

P1774 L6: model's -> models'

Done.

P1775 L9: remain -> remains

Done.

L10: model studies -> model-based studies

Done.

L14: most of these models can provide daily temps or even sub-daily temps. You should at least mention the restriction of the model results that are available from RCN.

Done. We change the sentence in L13-L14 with "... on TP. Data at higher than monthly temporal resolution are not stored by the models in the PCN archive. Therefore TSL diagnosis......"

P1776 L3: you can provide one supplementary plot/table to show that 38 m vs 3 m does not affect the MAGT method results for CLM.

Done. Please see our detailed reply to Question 3 of Referee #1 and Table 6.

P1779 L1: What is the reason to use MIROC-ESM here? As you say it is not comparable to offline-forced models. I don't see the input of mentioning that to this manuscript.

Sorry, this is misleading. We delete all mention of MIROC-ESM.

P1781 L3: lose the comma

Done.

L4: if you are talking about Cohen's paper, then you should put the reference out of parenthesis

Done.

L18: sites -> sites'

Done.

P1783 L24: I don’t understand what you are talking about, when you choose K>0.4, then all models except CLM passes for the MAAT method. And for the criteria K>0.2, UVic also passes for MAAT and F methods. Please clarify which methods you are talking about here.

This is a misunderstanding. We should keep in mind that the permafrost derived with TSL, MAGT and SFI are modeling results, but those of MAAT and F not. Thus here we only focus on K associated with TSL, MAGT and SFI to evaluate the modeling ability.

To avoid this misinterpretation, we added a sentence in P1783 L23, “......K (Sect. 3.3), and we limit discussion to the K associated with TSL, MAGT and SFI, which are calculated with simulated soil temperatures. If we take the (arbitrary) ......”

P1784 L3: Please mention which figure or table you are referring to. In which figure do we see the seasonal cycle amplitude of ISBA is better matched than others? In Fig 4d, ISBA results are not so similar to the observed in terms of amplitude. In Fig 4a and 4c, almost all models (except LPJ-GUESS) have good matching amplitudes. And in Fig 4b is the only plot where we can see a better match of ISBA. If this is the case, you only plot where we can see a better match of ISBA. If this is the case , you should revise this sentence. Yes in Table 4, we can see ISBA is the only one that satisfies the <2C condition for all sites/ depths but considering there is only 2 sites for the lower depth (2.63), it is hard to generalize

This is a misunderstanding. 1) all the numbers of Table 4 are calculated from Fig. 4. So we can just focus on Table 4 here. 2) We didn’t mean to emphasize ISBA is better than others on seasonal cycle amplitude simulation. “Bias is ≤ ±2.0 ˚ C” is a very loose threshold. Only one model (ISBA) seems to meet this criterion. The sentence is changed : ".. Only one model (ISBA) is consistent with the limited observations".

P1785 L14: CoLM model does not show lower mean annual temperatures than CLM or JULES according to Table 4

Yes you are right, we neglected to mention that we were referring to the selected region. We rephrase this part:

" We investigate both the air and ground temperature (Fig. 5) of the selected region (the region shown in Fig. 1), which is the coldest part of TP and should be permafrost. CoLM simulates no permafrost in the selected region despite CoLM having lower mean annual ground temperatures for the 3 m layer than many other models (ISBA, CLM4.5 and JULES) (Fig. 5). "

L18: classed -> classified

Done.

L18: permafrost -> non-permafrost?

No. “**precluding** it being classed as permafrost with the TSL method” means it is non-permafrost.

For the last paragraph of section 4.4, you should mention that you are talking about the selected region rather than the observational sites.

Done. We add a sentence in L13, “...variability (Table 3). We investigate both the air and ground temperature (Fig. 5) of the selected region (the region shown in Fig. 1), which is the coldest part of TP and should be permafrost. CoLM simulates no permafrost in the selected region despite CoLM having lower mean annual ground temperatures for the 3 m layer than many other models (ISBA, CLM4.5 and JULES) (Fig. 5). However, ...”

L21: revise the first sentence of sect 5. Too long to deliver the message clearly.

Done. It reads now, “As discussed in Sect. 4, the most noticeable ground temperature discrepancies among the 6 models are the underestimation of soil temperature by LPJ-GUESS and the overestimation of soil temperature by UVic, which lead to the largest biases in simulated permafrost area.”

P1788 L10-12: sentence is too long to make sense. Separate the last part starting with “observation-based Wang06...”

Done.

P1789 L2: give references to show the need for model improvements and model depths extensions.

The reference for model depths is mentioned below, see P1789 L3, “Nicolsky et al. (2007) recommend a soil column of at least 80 m for models applied to arctic and boreal regions.”

Section 4 contains both results and discussions. Put the title correctly or make a better separation between pure results and discussion points.

Done. We change the title and add a new discussion section 6.

You mention model soil depth could be a reason but you don’t discuss that in your discussion sections.

Done. This question is essentially the same as referee #1 question 3 to which we refer for our detailed answer. We address the point in the paper with a new Table 6 and discussion in a new section 6.2.

Fig 1: your legend is not clear. What is the “selected region”? It is only described later in the section 3.4. You should describe it also in the figure caption.

Done. It now reads:

“Figure 1. Permafrost maps derived from different diagnostic methods and models compared with Wang06 map. Permafrost inside the common modeling region is used for all-models inter-comparison, while permafrost outside allows further evaluation over the whole TP for CLM4.5, CoLM, JULES and UVic. The observation-based map of permafrost (Wang et al., 2006) is re-gridded to match model resolution. The selected area in the western TP (33°- 36°N, 82.5°- 85.5°E) is used to examine across-model differences in Figure 5. Insets show location map of TP and how the common region is related to the TP."

It doesn’t make sense to put Wang06 map in between methods. You should make a separation between methods and observational map.

The intent is to separate the direct and indirect methods. This also placed observations in the middle of the plot where it is most easy to compare across all the model simulations

Can you also put description to the smaller two maps under the panel (Tibet and common region).

We add a description to the caption: "We also show a location map of TP and how the common region is related to the TP."

Site locations are not very visible. Try to choose another marker and make them bolder

Done.

Fig 2: Can you explain how you calculated the error bars from resolution differences?

The error bar is calculated as half of the averaged grid cell area of the model, so is model resolution dependent. We add this clarification into the legend for Fig. 2.

And there was a mistake about the error bar of UVic. We correct it and update the Fig. 2 now.

Fig 4: Mention the reason of using only upper soil temp for D110 and only the subsoil temp of D105 sites in the caption

Done. It reads now:

“Monthly soil temperature variations at 3 stations from models and observations. (a) and (c) soil temperature of top layer. (b) and (d) soil temperature of deeper layer, 1996-2000. “Mean” denotes annual average temperature. We use the topmost available soil temperatures (0.04 m at D66 and D110, no good data for D105) and lowest available ones (2.63m at D66, 3m of D105), while D110 has only temperatures at 2 m depth.”

Fig 6: You should mention the source and description of observations in the figure caption. Explain OBS_0.5, OBS_clm4.5, OBS_uvic in the caption.

Done. It reads now:

“Figure 6. Winter snow depth for the common region, averaged over 1980–2000. Note the nonlinear color scale. We use the Long Time Series Snow Dataset of China (Che et al., 2008)(http://westdc.westgis.ac.cn) as observed snow depth. The observed snow depth plot is further interpolated onto the models’ resolutions as “OBS_”. The OBS_05 is in 0.5 °resolution for CoLM, ISBA, JULES and LPJ-GUESS. The OBS_CLM4.5 and OBS_UVic are in the resolutions of CLM4.5 and UVic separately.”

Table 5. Start all words with capital letter.

Done. It reads now “Table 5. Description of Model Characteristics Relevant to Soil Temperatures.”

What does “snow cover: none” mean for UVic? No snow representation? This has to be mentioned because it affects everything for soil thermal dynamics...

UVic has both snow fall and a snow scheme, but zero snow cover depth. On TP the snow is removed by sublimation, as we mentioned. Please see our detailed reply to your first comment. And we improve Fig.6 accordingly.

What do you mean by “unfrozen water effect during phase change”? Does that mean no freezing/ thawing occurs in CoLM, LPJ-GUESS, and UVic?

We clarify this in section 3.1" That is all models calculate soil thermal properties as a function of soil moisture and consider the phase change of water/ice, but CoLM and LPJ-GUESS do not consider transformation to ice of water solute mixtures below 0 ℃."

Diagnostic and model dependent uncertainty of simulated Tibetan permafrost area

Wenli Wang[1], Annette Rinke[1,2], John C. Moore[1], Xuefeng Cui[1*], DuoyingJi[1], Qian Li[3], Ningning Zhang[3], Chenghai Wang[4], Shiqiang Zhang[5], David M. Lawrence[6], A. David McGuire[7], Wenxin Zhang[8], Christine Delire[9], Charles Koven[10], Kazuyuki Saito[11], Andrew MacDougall[12], Eleanor Burke[13], Bertrand Decharme[9]

[1]{State Key Laboratory of Earth Surface Processes and Resource Ecology, College of Global Change and Earth System Science, Beijing Normal University, Beijing 100875, China}

[2]{Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Potsdam, Germany}

[3]{Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China}

[4]{School of Atmospheric Sciences, Lanzhou University, Lanzhou, China}

[5]{College of Urban and Environmental Sciences, Northwest University, Xi' an, China}

[6]{NCAR, Boulder, USA}

[7]{U.S. Geological Survey, Alaska Cooperative Fish and Wildlife Research Unit, University of Alaska, Fairbanks, USA}

[8]{Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden}

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

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

[11] {Department of Integrated Climate Change Projection Research, Japan Agency for Marine-Earth Science and Technology, Yokohama, Kanagawa, Japan}~~{Research Institute for Global Change, Japan Agency for Marine-Earth Science and Technology, Yokohama, Kanagawa, Japan~~

[12]{School of Earth and Ocean Sciences, University of Victoria, Victoria, BC, Canada}

[13]{Met Office Hadley Centre, Exeter, UK}

Correspondence to: Xuefeng Cui (xuefeng.cui@bnu.edu.cn)

Abstract

We perform a land surface model intercomparison to investigate how the simulation of permafrost area on the Tibetan Plateau (TP) varies ~~between~~ among 6 modern stand-alone land surface models (CLM4.5, CoLM, ISBA, JULES, LPJ-GUESS, UVic). We also examine the variability in simulated permafrost area and distribution introduced by 5 different methods of diagnosing permafrost (from modeled monthly ground temperature, mean annual ground and air temperatures, air and surface frost indexes). There is good agreement (99 to 135 $\times 10^4$ km^2) between the two diagnostic methods based on air temperature which are also consistent with the ~~best current~~ observation-based estimate of actual permafrost area (101 $\times 10^4$ km^2). However the uncertainty (1 to 128 $\times 10^4$ km^2) using the three methods that require simulation of ground temperature is much greater. Moreover simulated permafrost distribution on TP is generally only fair to poor for these three methods (diagnosis of permafrost from monthly, and mean annual ground temperature, and surface frost index), while permafrost distribution using air temperature based methods is generally good. Model evaluation at field sites highlights specific problems in process simulations likely related to soil texture specification, vegetation types and snow cover. Models are particularly poor at simulating permafrost distribution using the definition that soil

temperature remains at or below $0^{o}C$ for 24 consecutive months, which requires reliable simulation of both mean annual ground temperatures and seasonal cycle, and hence is relatively demanding. Although models can produce better permafrost maps using mean annual ground temperature and surface frost index, analysis of simulated soil temperature profiles reveals substantial biases. The current generation of land surface models need to reduce biases in simulated soil temperature profiles before reliable contemporary permafrost maps and predictions of changes in permafrost distribution can be made for the Tibetan Plateau.

1 Introduction

The Tibetan Plateau (TP) has the highest and largest low-latitude frozen ground in the world, with more than 50% of its area occupied by permafrost (Zhou et al., 2000). The unique geography and plateau climate make~~s~~ the permafrost on TP very different from the Arctic. The TP permafrost is warmer, with only discontinuous and sporadic permafrost (Zhou et al., 2000), has less underground ice (Ran et al., 2012), and has no large forests (Wu, 1980). The active layer thickness ranges from 1 m to 3 m, with some intensely degraded area reaching 4.5 m (Wu and Liu, 2004; Wu and Zhang, 2010; Zhang and Wu, 2012). Freeze/thaw cycles, and the extent of permafrost play~~s~~ an important role in the thermal state of TP. The underlying surface temperature contrast between TP and Indian Ocean is an important controlling factor for both the Asian monsoon~~,~~ and the wider general atmospheric circulation (Xin et al., 2012). As TP gets intensely warmer (IPCC, 2013; Wu et al., 2013), the impact of degraded permafrost on desertification (Li et al., 2014; Yang et al., 2010; Li et al., 2005), water cycling (Cheng and Jin, 2013; Yao et al., 2013), carbon budget (Dörfer et al., 2013; Wang et al., 2008; Schuur et al., 2008;), and infrastructure (Wu and Niu, 2013; Yu et al., 2013) has also become active research topics.

Hence, the simulation of TP permafrost is motivated both by its global importance and by its unique properties. A number of land surface models (LSMs) (e.g., CLM4.0, CoLM, SHAW, Couple Model~~,~~ and FSM ~~and VIC~~) have been applied at individual station locations on TP to reproduce soil thermo-hydro dynamics (Li et al., 2009; Wang and Shi, 2007; Xiong et al., 2014; Zhang et al., 2012). Simulations of ground temperature and moisture variations are relatively realistic when using observed atmospheric forcing (Guo and Yang, 2010; Luo et al., 2008). The results were improved by setting appropriate permafrost parameters for soil organic matter contents and soil texture properties (Luo et al., 2008; Wang et al., 2007; Xiong et al., 2014). CLM4.0 has also been used to provide future projections of permafrost extent for the whole TP (Guo and Wang, 2013; Guo et al., 2012), and simulates 81% loss of permafrost area by the end of 21st century under the A1B greenhouse gas emissions scenario. This raises the question of how reliable the estimate is in comparison with results from other models.

Simulations of Northern Hemisphere (NH) permafrost area showed large differences amongst Coupled Model Inter-comparison Project (CMIP5) models (Koven et al., 2013; Slater and Lawrence, 2013). Moreover, different diagnostic methods, using either a direct method, which relies on model simulated ground temperatures, or indirect methods inferred from air temperatures and snow characteristics also lead to quite different permafrost areas. Slater and Lawrence (2013) applied two direct methods to nineteen CMIP5 models and found differences of up to 12.6×10^6 km^2 in diagnosed NH permafrost area. Saito (2013) showed that differences in pre-industrial NH continuous permafrost area between direct and indirect methods were around 3×10^6 km^2. This raises the question why different methods arrive at different estimates and which method is better suited.

A reliable simulation of permafrost extent is important, since permafrost is a comprehensive reflection of soil thermo-hydro dynamics that is hard to measure directly except at sparse observational sites. Further, reliable present-day simulations can contribute to an increased confidence in simulations of future permafrost degradation by these models. ~~We note that this approach provides information on the ability of models to simulate permafrost in a region that is both warmer and physically different from where they were "tuned"~~ We note that this approach provides information on the ability of models on the warmer and physically unique TP permafrost in a NH simulation, hence providing some test of reliability for simulations of present and future global permafrost over TP.

To date, an examination of the uncertainties in model-derived TP permafrost area has not been attempted. One way of estimating this uncertainty is to explore a single model and to perform a set of sensitivity experiments in which the model parameters are modified (e.g., Dankers et al., 2011; Essery et al., 2013; Gubler et al., 2013). An alternative approach is to explore an ensemble of multiple models where the uncertainty is discussed in terms of the spread among the models (e.g., Koven et al., 2013; Slater and Lawrence, 2013). Here we follow the second approach and examine the uncertainty of TP permafrost simulations by an ensemble of 6 state-of-the-art stand-alone land-surface schemes. The models are from the Permafrost Carbon ~~Research~~ Network (~~Permafrost-R~~PCN; http://www.permafrostcarbon.org/) and include a broad variety of snow and ground parameters and descriptions, along with a clear experimental design under prescribed observation-based atmospheric forcing. The first focus of our paper is therefore the quantification of the uncertainty in the simulated TP permafrost area due to the ~~model's~~ models' structural and parametric differences. Further, using time series of soil temperature from the few available TP stations, we discuss the biases in relation to the land surface model description (e.g. soil texture, vegetation and snow cover). We also discuss in the paper the uncertainty due to the different methods to diagnose the TP permafrost area, with 5 different (direct and indirect) methods.

In section 2 we introduce the different methods used to derive permafrost extent for the TP from LSMs. Section 3 describes the applied model data, the observation-based estimate of TP permafrost map, the method to assess the agreement of simulated versus observation- based estimate of permafrost maps permafrost maps, and ground temperature data to evaluate soil thermal profiles simulated by the models. Results and discussion are presented in sections 4 and 5, and conclusions are summarized in section 6.

2 Permafrost Diagnosis

We make use of all five major permafrost diagnostic methods promoted in the literature (Slater and Lawrence, 2013; Guo et al., 2012; Guo and Wang, 2013; Wang et al., 2006; Wang, 2010; Nan et al., 2002; Nan et al., 2012; Saito, 2013; Ran et al., 2012; Wang et al., 2006; Jin et al., 2007; Xu et al., 2001; Nelson and Outcalt, 1987). Since the model intercomparison relies on LSMs that are all driven at monthly resolution, the methods we use are tailored, as usual, to reflect the forcing data resolution. The model-derived TP permafrost maps are shown in Figure 1. The modeling spatial domain is not consistent among the models. CLM4.5, CoLM, JULES and UVic cover the whole TP while others (ISBA, LPJ-GUESS) do not (Table 1). We mainly focus on the common modeling region (Figure 1) to discuss differences between models and methods, but also give the results for whole TP for the four models that produce them.

In detail, the five methods are:

1) Temperature in Soil Layers (TSL)

The TSL method allows a direct diagnosis of permafrost from modeled soil temperature

(Slater and Lawrence, 2013). The standard definition of permafrost is that ground remains at or below 0 ℃ for at least two consecutive years. Many recent modeling studies (e.g., Guo et al., 2012; Guo and Wang, 2013; Slater and Lawrence, 2013 and references therein), have consistently adapted this for land surface and earth system models by defining a model grid cell as permafrost if the simulated ground temperature (of at least one level in the upper soil) remains at or below 0 ℃ for at least 24 consecutive months. Furthermore, these model-based studies are limited by the maximum soil depth of the models (Table 1). Hence, we analyze the ground temperatures down to a depth of 3 m, which should be satisfactory as this range spans the observed active layer thickness on TP. ~~Since the models do not provide ground temperatures at a higher temporal resolution than the monthly time scale, the~~ Data at higher than monthly temporal resolution are not stored by the models in the PCN archive. Therefore TSL diagnosis is calculated from monthly mean soil temperatures, which has been previously demonstrated to be a viable substitute for model-based estimates of permafrost both on TP (Guo et al., 2012; Guo and Wang, 2013), and for the Arctic (Slater and Lawrence, 2013).

2) Mean Annual Ground Temperature (MAGT)

Permafrost is detected if the mean annual ground temperature at the depth of zero annual amplitude is at or below 0 ℃ (Slater and Lawrence, 2013). Some papers use a slightly higher critical temperature, e.g. 0.5 ℃ (Wang et al., 2006; Wang, 2010; Nan et al., 2002), which has been found to fit TP observations well. Slater and Lawrence (2013) suggested MAGT as an indicator of deeper permafrost. The problem with this definition is that many models have quite shallow soil depth (Table 1), and of course, zero amplitude would require great (actually infinite in steady state) soil depth. For practical purposes, we use MAGT at 3 m depth (the approximate base of the active layer) and the common critical temperature of 0 ℃. Although annual ground temperature amplitudes at 3 m depth are still several degrees, they are much smaller than the amplitudes in upper layers (section 4.3). We investigated one model with a larger depth range (CLM4.5; Table 1) in more detail, but found that the results using MAGT at 38 m depth do not significantly change the derived permafrost area.

3) Surface frost index (SFI)

Originally, Nelson and Outcalt (1987) introduced the surface frost index SFI*, also used in Slater and Lawrence (2013):

$$SFI^* = \frac{\sqrt{DDF_a^*}}{\sqrt{DDF_a^*} + \sqrt{DDT_a}} \qquad (1),$$

Where DDF_a^* and DDT_a are the annual freezing and thawing degree-day sums, both calculated using air temperature (indicated by *a* subscripts), and with DDF_a^* further modified to correct for the insulating effect of snow cover (indicated by the *superscript). In this way, SFI^* is designed to reflect the ground surface thermal conditions by combining snow insulation effect with air temperature. However, the snow insulation effect alone can not account for the soil structure complexity. So here we calculate surface frost index directly from the ground surface temperature (indicated by *s* subscripts) (Nan et al., 2012), using an asymmetric sinusoidal annual temperature cycle fitted to the warmest and coldest monthly temperatures ($\overline{T_h}$,$\overline{T_c}$) and a frost angle (β) (Nan et al., 2012):

$$SFI = \frac{\sqrt{DDF_s}}{\sqrt{DDF_s} + \sqrt{DDT_s}} = \frac{1}{1 + \sqrt{\dfrac{\beta(\overline{T_h} + \overline{T_c}) + (\overline{T_h} - \overline{T_c})\sin\beta}{(\beta - \pi)(\overline{T_h} + \overline{T_c}) + (\overline{T_h} - \overline{T_c})\sin\beta}}} \qquad (2),$$

Nan et al. (2012) report good results using this surface frost index on TP with values of SFI ≥ 0.5 to indicate permafrost.

4) Air frost index (F)

Nelson (1987) calculated F from an equation analogous to (2), but using monthly air temperature rather than ground surface temperatures. Where F ≥ 0.5 defines permafrost. We follow suit and use F to assess the effects of air temperature forcing. Although many authors have criticized F as a permafrost indicator, F has been used in recent work, though in modified forms. For example, Saito (2013) calculated mean annual air temperature (MAAT) as $MAAT = (DDT_a - DDF_a)/365$, where DDT_a and DDF_a, are thawing index and freezing index as defined earlier which means that MAAT in Saito (2013) is a proxy for F.

5) Mean Annual Air Temperature (MAAT)

A critical value of MAAT is often used to derive the southern boundary of permafrost (Ran et al., 2012; Wang et al., 2006; Jin et al., 2007). The -2 ℃ isotherm of MAAT has been found to fit well with TP observation- based permafrost maps (Xu et al., 2001). MAAT has been used to compare the air temperature based permafrost area with permafrost areas derived by other methods (Koven et al., 2013; Saito et al., 2013). Note that the calculation method of MAAT in Saito et al. (2013) is slightly different from that used in other works. Here we calculated MAAT traditionally, as the average of 12 monthly 2 m air temperatures.

All the 5 diagnostic methods are summarized in Table 2. The three direct methods (TSL, MAGT, SFI) are based on simulated ground temperatures, while the two indirect methods (F and MAAT) use the prescribed air temperature. SFI is mainly controlled by air temperature and snow cover, but it also depends on how the soil is parameterized, so SFI is somewhat closer to the indirect methods than are TSL and MAGT.

The 3 methods introduced in the 1980s (SFI, F, MAAT), were designed to map permafrost based on the assumption that the permafrost distribution is related to climatic parameters. Although permafrost processes are directly represented in climate models nowadays, the simulated soil temperatures have considerable errors, and the directly diagnosed permafrost area has model-dependent biases (Koven et al., 2013; Slater and Lawrence, 2013). Therefore the older indirect diagnostic methods are also still very commonly used (e.g., Wang et al., 2006; Jin et al., 2007; Ran et al., 2012; Nan et al., 2012; Slater and Lawrence, 2013; Saito, 2013; Koven et al., 2013). TP permafrost area directly diagnosed from the simulated monthly soil temperatures (TSL) is not superior to the other methods in comparison with the observation-derived permafrost map (Figures 1 and 2). Hence, we consider all the 5 diagnostic methods to quantify the full range of uncertainty in the model-derived permafrost maps.

Since the forcing air temperatures of LSMs were not the same due to discrepancies in the historical temperature (and precipitation and other forcing fields) datasets used by the individual models (Table 1), we use the indirect methods to quantify forcing differences. If these differences are not too large, we can attribute the differences in the direct method-derived permafrost areas primarily to differences of modeled land surface processes. Across-model and across-method variability is listed in Table 3. As we use fairly small numbers of methods and models, rather than defining uncertainty in terms of ~~SD~~standard deviation, we choose to use the full range of values from the simulations and define uncertainty as maximum-minimum values among the models.

3 Data and Analysis Approach

3.1 Data from stand-alone LSMs

Output from six stand-alone LSMs participating in the inter-model comparison project "Vulnerability of Permafrost Carbon to Climate Change~~Research Coordination~~

~~Network (RCN-Permafrost)~~" (http://www.permafrostcarbon.org/) is analyzed in this study (Table 1). The simulations have been generally conducted for recent decades from 1960 to 2009 using monthly resolution climate forcing input data. Each modeling team was free to choose appropriate driving data sets for climate, atmospheric CO_2, N deposition, disturbance, soil texture, etc., as used in their standard modeling system. Model spin-ups are also different, but they are long enough (around 1 000 years) to ensure that the deep carbon is in equilibrium. The LSMs use different horizontal model resolutions and different soil layer divisions (Table 1). ~~We also analyzed (but do not show here) output from a coupled earth-system model (Miroc-ESM). In contrast with the land surface models we present here, the Miroc-ESM coupled model generates its own air temperatures, which over TP, were 4-8 ℃ cooler than temperatures in the other driving datasets. This creates issues with other model fields such as snow thickness and albedo which make comparison of permafrost processes more difficult than with the stand-alone LSMs.~~

Our analysis is based on monthly averages of the driving air temperature and simulated ground temperature. As three models (CoLM, JULES and LPJ-GUESS; Table 1) have shallow soil layers, we restrict our analysis to the common depth range spanning near surface to 3 m. Ground temperatures were linearly interpolated onto the common depths: 0.05, 0.1, 0.2, 0.5, 1, 2, 3 m. Since there is no ground surface temperature output, we linearly extrapolate the ~~below~~ top two layers' soil temperatures onto the ground surface. For CLM4.5, CoLM, ISBA and LPJ-GUESS, the first layer soil depth is no deeper than 0.01 m and the second layer soil depth is no deeper than 0.05 m. For JULES and UVic, the first layer soil depth is 0.05 m and the second layer soil depth is no deeper than 0.18 m. Most TP permafrost work has been post-1980 (Guo and Wang, 2013; Nan et al., 2012), so we choose 1980 as the start of the analysis period. The end is limited to the year 2000 by results from the JULES model (Table 1).

The LSMs in this study considered the following processes: dynamic vegetation, carbon cycling (Rawlins et al., 2015), snow, near-surface hydrological budget, soil thermal dynamics (Peng et al., 2015) and the treatment of freezing soil. Sophistication in the treatment of these processes varies amongst the models with each having specific parameterizations, In this study we investigate some key schemes and parameters that are important for permafrost simulation: 1) Unfrozen water / phase change. All models calculate soil thermal properties as a function of soil moisture and consider the phase change of water/ice, but CoLM and LPJ-GUESS do not consider transformation to ice of water solute mixtures below 0 ℃, which is a key feature in soil freezing and thawing. 2) Surface organic layer insulation. Only CLM4.5 and ISBA consider the insulating effect of moss. 3) Soil texture parameterization. The specified fraction of clay and sand in soil differs. LPJ-GUESS specifies the same soil texture for the TP as for the Arctic. 4) Organic soil fraction treatment. The organic content of soil differs among the models. LPJ-GUESS sets the same value for TP as for the more organically rich permafrost of the Arctic. 5) Snow processes. ISBA, LPJ-GUESS and UVic set static snow layers. UVic uses an implicit snow scheme while LPJ-GUESS uses the Bulk-layer scheme, which are both simpler than the dynamic multi-layer snow scheme of some other land models.

3.2 TP permafrost observation-based map

Mapping permafrost on TP is challenging due to absence of field observations, especially in the central and western parts where permafrost is widespread. In practice, permafrost maps on TP have been statistical models based on a compilation of earlier maps, aerial photographs, Landsat images and terrain analysis (Ran et al., 2012; Shi et al., 1988; Li and Cheng, 1996; Nan et al., 2002) as well as on some MAGT and MAAT data from the few long-term monitoring sites (Ran et al., 2012; Wang et al., 2006). The classification and therefore the mapping of TP permafrost is not consistent across the different studies (Ran et al., 2012). Thus there is a large spread of observation-based TP permafrost area estimates from 110×10^4 km^2 (Wang et al., 2006) to 150×10^4 km^2 (Shi and Mi, 1988; Li and Cheng, 1996).

The mostly widely used map by Li and Cheng (1996) has large differences from other maps, and shows excess permafrost in the southeast where permafrost can only exist on extremely cold mountains (Gruber, 2012). The International Permafrost Association (IPA) map (Brown et al., 1997; Heginbottom, 2002) is the most widely used in NH permafrost analysis. However, the IPA map is not well suited for TP because the data and information in this map is based on the map made by Shi et al. (1988) which has not been updated since.

We use the 1 : 4,000,000 Map of the Glaciers, Frozen Ground and Deserts in China (Wang et al., 2006, hereafter refered to as the “Wang06 map”) as the primary reference. The map is based on MAGT (Nan et al., 2002) with 0.5 ℃ as the boundary between permafrost and seasonally frozen ground. Nan (2002) fitted a multiple linear regression between latitude, altitude and MAGT, from all 76 TP stations having borehole data, and extrapolated this regression to the whole TP with a 1 km resolution DEM to get the MAGT distribution. The Wang06 map was re-gridded to match the different model resolutions and spatial domain (see ~~observation~~“Wang06 map” column in Figure 1), and the different permafrost areas derived from the methods and models are compared with the Wang06 map in Figure 2.

We emphasize that the Wang06 map is subject to uncertainty as it is based on a relatively sparse set of observations and then statistical extrapolation. Nan et al. (2013) pointed out that permafrost was overestimated in the western TP in both the maps by Li and Cheng (1996) and Wang et al. (2006). However, a better permafrost map covering the whole TP is not available.

3.3 Measure of agreement between simulated and Wang06 permafrost maps

To evaluate the agreement of simulated permafrost map with the Wang06 map, we calculate the Kappa coefficient (Cohen, 1960; Monserud and Leemans, 1992; Wang, 2010), *K*, which measures the degree of agreement between two maps.

$$K = \frac{\left(s/n - (a_1 b_1 + a_0 b_0)/n^2 \right)}{\left(1 - (a_1 b_1 + a_0 b_0)/n^2 \right)} \qquad (3)$$

Where the total number of the map points is n, and s is the number of points where simulation and observational estimate agree. The numbers of Wang06 map cells with permafrost is a_1, and those without are a_0, and the corresponding simulated map cell numbers are b_1 and b_0. The calculated *K* matrix of simulated and Wang06 permafrost maps is presented in Figure 3. Empirically~~,~~ and statistically arbitrary quality values for *K* have been proposed, ~~(~~e.g. Cohen~~,~~ (1960)~~, who~~ suggested that $K \geq 0.8$ signifies excellent agreement, $0.6 \leq K < 0.8$ represents substantial agreement, $0.4 \leq K < 0.6$ represents moderate agreement, $0.2 \leq K < 0.4$ represents fair agreement, while lack of agreement corresponds to $K < 0.2$. There is a sample size issue in estimating the confidence of *K* and this can be a factor when very small numbers of grid points are available (here this applies to UVic).

3.4 Data used to examine model thermal structures

The derived permafrost maps depend on the modeled ground thermal structures. However, field studies on TP are quite limited, and we have only short duration (1996-2000) ground temperature profiles obtained from the GEWEX Asian Monsoon Experiment (GAME)-Tibet (Yang et al., 2003) at three permafrost stations (D66, D105, D110; Figure 1) in the central TP to compare with model results. The three stations are located along the Qinghai-Tibet Highway. D66 station is in the front edge of alluvial fan, with almost no vegetation. The soil is mainly composed of gravels, sands and pebbles. D110 is in the southern bank of ZhaJiaZangBu River. The ground is a wetland covered with short-stature emergent vegetation. The upper layer soil is composed of coarse and fine sand. The lower soil layer is mainly composed of fine sand. D105 is in the northern side of the Tanggula Mountain range. The ground surface is relatively flat, covered with plateau meadow. The soil is composed of both coarse and fine sand. The vertical profile of observed soil temperature of D66 extends from 0.04 m to 2.63 m, of D110 from 0.04 m to 1.8 m, and of D105 from 0 to 3 m. However the data continuity of the top layer temperature in D105 is not good. To examine modeled ground temperatures, we~~We~~ present the top (0.04 m) and deeper (2.63 m or 3 m) soil layer temperatures (modeled temperatures were weighted bi-linear interpolated onto the station locations) in Figure 4 and Table 4. We also give a short description of the sites vegetation and soil texture information, both from observation and models.

We also analyze monthly air and ground temperatures in a selected area in the western TP (33° - 36° N, 82.5° - 85.5° E, Figure 1) to examine across-model differences (Figure 5). The air temperature is also different among the models, especially in winter season, though the differences are much smaller than soil temperatures differences. As this region is the coldest part of TP (according to the annual mean air temperature) the permafrost is widely distributed, and the active layer thickness is less than 3 m. However, TSL method derived permafrost areas vary significantly among the models in this area (Figure 1). Despite the lack of any ground temperature observations in this area, the definite presence of permafrost makes it useful to look at the ground thermal structure of each model as well as their differences as a means of interpreting the calculated permafrost areas.

4 Results and Discussion

4.1 Uncertainties in air-temperature-derived permafrost area

Air temperature–derived permafrost maps are investigated with the two indirect methods, F and MAAT. Figures 1 and 2 compare both Wang06 and model-derived permafrost maps, and show that F produces consistently excessive permafrost area compared with MAAT. That is because the empirical threshold of -2 ℃ for MAAT fits well with TP observations (Xu et al., 2001), while F ≥ 0.5 is a theoretical assumption, which has been reported to overestimate permafrost area (Nelson and Outcalt, 1987; Slater and Lawrence, 2013). Accordingly, Figure 3 shows that F-derived permafrost is less consistent with ~~observation~~ Wang06 map (model average K = 0.3 for the common region) than MAAT-derived permafrost area (K = 0.5).

Across-model variability (Table 3) for the MAAT-based method is 14×10^4 km^2 and for the F-based method is 17×10^4 km^2, equivalent to about 14 % ~ 17 % of the Wang06 permafrost area inside the common modeling region (101×10^4 km^2). This variability is much smaller than the 56% calculated by Slater and Lawrence (2013) for the CMIP5 models with the SFI* method for NH permafrost area. The relatively smaller difference among the models here is because, although the temperature forcing was not identical among models, the mean annual air temperature and its spatial variability in the permafrost region are quite similar (between -6 ℃ and -8 ℃). ~~Hence most of the differences among the indirect methods that use air temperature to derive permafrost area can be attributed to different model horizontal resolutions.~~ Since the differences in permafrost extent using the air temperature based indirect methods are relatively small, the differences in the direct method derived extents can primarily be attributed to the LSMs structural and parametric differences.

4.2 Uncertainties in model–derived permafrost area

There is a large across-model variability of permafrost area derived from direct methods (TSL, MAGT and SFI) (Figures 1, 2; 111~120 $\times$ 10^4 km^2; Table 3) and it is similar for all the 3 diagnosis methods. This across-model variability is much larger than the variability using the indirect methods discussed in Section 4.1, and is equivalent to 110-112% of Wang06 permafrost area for the common modeling region. CMIP5 across-model variability derived from TSL in NH permafrost area was similarly large (Slater and Lawrence, 2013; Koven 2013). Clearly this points to large across-model differences in ground thermal structures.

The across-method (TSL, MAGT and SFI) variability in permafrost area (Figures 1, 2; Table 3) is very variable between models: UVic and LPJ-GUESS have smallest ranges (up to 9 x 10^4 km^2), while CoLM has the largest (87 $\times$ 10^4 km^2) (Table 3), near to the total permafrost area of the common region. Thus the across-direct method range is similar to the across-model range. Slater and Lawrence (2013) also emphasized the variable across-method variability for NH permafrost area between models, however Saito (2013) showed insignificant variability across both direct and indirect methods for derived pre-industrial NH continuous permafrost area.

4.3 Model evaluation based on K and ground temperature profile

A good land surface model should adequately simulate the seasonal and annual ground temperature profiles. Hence one quality test for a model is that it should be able to produce 'good' permafrost maps, which we define as agreement with the observation-based map, based on all the three direct diagnostic methods. The applied criterion is the kappa coefficient K (section 3.3)~~.~~, and we limit discussion to the K associated with TSL, MAGT and SFI, which are calculated with simulated soil temperatures. If we take the (arbitrary) threshold K– ≥ 0.4 (indicating "moderate agreement”), then no model passes this test for the common simulation region, while reducing the threshold to K ≥ 0.2 (“fair agreement”) allows most models and methods to pass while UVic stands out as a clear failure (Figure 3).

If the criterion for acceptable model bias is ≤ ±2.0 ℃, then simulations of mean annual ground temperatures from most models (CLM4.5, CoLM, ISBA and JULES) agree with the observations, but only the simulation of seasonal cycle amplitude of one model (ISBA) ~~agrees with observations~~ is consistent with the limited observations. However, if the criterion is bias ≤ ±1.0 ℃, then no model agrees with observations for neither mean annual ground temperature nor the seasonal cycle amplitude (Figure 4, Table 4).

We now look at the performance of the 2 models with larger biases in mean annual ground temperature: LPJ-GUESS and UVic. LPJ-GUESS simulated too cold (by more than 3 ℃) mean annual ground temperatures for both the surface and deeper layers (Figure 4, Table 4). The summer temperatures simulated by the model in the surface layers are especially cold, with maximum temperatures lower than observation by 8 ℃ (Figures 4a, c) and its ground temperature amplitude is substantially underestimated (Table 4), which must greatly limit the summer thaw depth. This cold soil results in substantial overestimation of permafrost area (119 ~ 131×10^4 km^2; Table 3, Figure 2) with small across-method variability.

UVic simulates a soil thermal state that is the warmest among the models, with the simulated mean annual ground temperature at D66 surpassing observation by more than 7 ℃ (Figure 4, Table 4). If the observational sites are representative then the generally too warm ground temperature in UVic is the reason for the extremely small simulated permafrost area (8×10^4 km^2; Table 3, Figure 2) with all direct methods, and hence to no across-method variability, and poor agreement with the Wang06 permafrost map (K < 0.1; Figure 3).

4.4 Method comparison based on K and ground temperature profile

Permafrost maps derived using MAGT and SFI often show larger area than TSL (Figure 2), with generally better agreement with the Wang06 map (Figure 3). The MAGT method simply defines a grid as permafrost as long as its 3 m mean annual ground temperature is colder than 0 ℃, and a permafrost threshold value of SFI ≥ 0.5 also only requires the mean annual ground surface temperature is lower than 0 ℃ (Nan, 2012). Figure 4 and Figure 5 show most models meet these criteria. However, assuming that the site observations are representative, the simulated mean annual ground temperatures of both surface and deeper soil layers often have obvious biases (≥ ±1 ℃) in all the models (Figure 4 and Table 4).

In general, model-derived permafrost distribution using the TSL method shows little agreement with the Wang06 map (Figures 1 - 3). In contrast with MAGT and SFI methods, the TSL method requires adequate simulation of both mean annual ground temperature and the seasonal cycle at monthly resolution (Figure 4, Table 4). This means that the TSL method is more susceptible to model errors, but it offers a more comprehensive insight into land model processes. CoLM is an extreme example of how a simulated permafrost map can be totally incorrect due to small errors in seasonal ground temperature. CoLM simulates nearly no TSL -derived permafrost (Figures 1, 2), accounting for much of the large across-model and across-method variability (Table 3). We investigate both the air and ground temperature (Figure 5) of the selected region (the region shown in Figure 1), which is the coldest part of TP and should be permafrost. CoLM simulates no permafrost in the selected region despite CoLM having lower mean annual ground temperatures for the 3 m layer than many other models (ISBA, CLM4.5 and JULES) (Figure 5).~~This is despite CoLM having lower mean annual~~

ground temperatures for the 3 m layer than many other models (ISBA, CLM4.5 and JULES). However, CoLMit simulates a larger seasonal amplitude than CLM4.5 and ISBA (Figure 5), so that, in the western TP, the monthly maximum 3 m ground temperatures in CoLM always surpasses 0 ℃ by around 0.2 ℃ (Figure 5c) precluding it being classifiedclassed as permafrost with the TSL method.

5 Discussion of the related mMain processes causing ground temperature discrepancies

As discussed in Sect. 4, the most noticeable ground temperature discrepancies among the 6 models are the underestimation of soil temperature by LPJ-GUESS and the overestimation of soil temperature by UVic, which lead to the largest biases in simulated permafrost area.In comparison with site observations, the most noticeable ground temperature discrepancies of the 6 models discussed in Section 4 and relevant for the most biased simulated permafrost area are the underestimation of soil temperature by LPJ-GUESS and the overestimation of soil temperature by UVic. There are many other, rather subtle, potential model discrepancies that we do not investigate in detail here. One example is the overestimation of the amplitude of the seasonal temperature cycle at deep depths in several models (Figures 4b and 4d; Table 4). Table 4 also shows that the observed vegetation and soil texture are mis-matched by all the models at each of the stations. Although it is a common problem to compare grid cell results against site data, model description of vegetation and soil texture is too simplified.

To help elucidate the causes of ground temperature discrepancies associated with soil processes we also inspect snow depth and vertical ground temperature gradients. We use the Long Time Series Snow Dataset of China (Che et al., 2008) (http://westdc.westgis.ac.cn) to examine the modeled snow depth. The complete dataset is composed of SMMR (1978-1987), SSM/I (1987-2008) and AMSR-E (2002-2010). According to Wang et al. (2013), the snow depth pattern and the significant seasonal snow characteristics of the satellite data are consistent with those of station data in most of our common TP region. The satellite data are different from station data on the southeast of TP (Wang et al., 2013), however, our analyzed common region does not include this part of TP. Thus this satellite data is reliable in this study. Here we use the data of SMMR and SSM/I to produce the winter (DJF) climatological distribution of 1980-2000 (Figure 6). Furthermore, we follow Koven et al. (2013) and calculated two vertical gradients to isolate processes: from the atmosphere to ground surface (Figure 7) and from ground surface to deeper soil (at 1 m depth) (Figure 8). While the first one is mainly controlled by the snow insulation, the latter is mainly determined by soil hydrology, latent heat and thermal properties. Important factors that influence the ground thermal structure are compared in Table 5. Since several models produce incomplete or not directly comparable output, we restrict ourselves to a qualitative assessment here.

The LPJ-GUESS simulated underestimation of soil temperature is not caused by a bias in the surface air temperature forcing (Figure 5, Table 4). Instead, this bias may be due to many factors such as inappropriate prescriptions of soil thermal properties, poor representation of soil hydrology, ~~and~~ mis-match of vegetation types, and weak coupling of soil water and vegetation cover. Figure 8 shows that the soil temperatures increase with depth, but LPJ-GUESS has a much smaller temperature gradient between the surface and the 1 m deep soil (0-2 K) than the other models. This suggests a different (larger) winter soil thermal conductivity probably associated with a high soil porosity and water content. LPJ-GUESS specifies the same soil texture for the TP as for the Arctic, which is mostly clay-like (Table 4). Clay has high water retention capacity. Many studies have reported that the soil on TP is immature, with coarser particles than typical for Arctic permafrost and with much less organic matter. Inappropriate soil texture classification will affect the simulated ground thermal structure. LPJ-GUESS underestimates the surface and top soil temperatures particularly in summer (Figures 4a, c, 5). Precipitation and hydrological processes determine the vertical profile of soil water content which can change the fraction of water and ice retained in different soil layers and influence soil thermal conduction. The energy required to melt the high water (ice) content in the surface soil layers in summer appears to lead to underestimated low summer temperatures compared with other models, and a phase lag in summer warming (Figures 4a and 4c).

In addition, LPJ-GUESS shows a similarly thick snow depth in the western part of Tibetan Plateau as CLM4.5 and CoLM (Figure 6), but does not show as large surface a temperature offset as those two models (Figure 7). That is because LPJ-GUESS has a fixed snow density (362 kg/m^3) which is higher than used in other models, and a relatively simple Bulk-layer snow scheme, with one static snow layer, unlike the dynamic multi-layer snow scheme of CLM4.5 and CoLM (Table 5).

UVic uses the same climate forcing as CLM4.5 (Table 1), but simulates much warmer ground temperatures than other models. In contrast with the other models, UVic has no snow cover in winter (Figure 6), which is consistent with grid cell surface albedo being year-round at values between 0.15-0.35. The simulated snow depth is derived from the prescribed winter precipitation, and the model's snow, energy and water balances. The lack of snow over TP in UVic likely indicates removal by sublimation. A too low snow albedo makes the snow gain energy that is lost through sublimation. Since it takes more energy to sublimate snow than it does to melt it, the latent heat flux should be, and is (not shown) higher in UVic than other models. However, despite the apparent snow sublimation - which should cool the soil, the ground surface temperatures in UVic are warmer than in all the models. The large absorption of short wave radiation allowed by the year-round low albedo provides this heat and is sufficient for there to be very little permafrost simulated by UVic for the TP.

ISBA, and especially JULES stand out from other models in their calculated winter temperature offsets: ground surface temperatures are colder than the driving air temperatures over much of the simulated region (Figure 7). Snow (Figure 6) and vegetation cover would normally be expected to~~should~~ provide insulation, making soil warmer than air temperatures in winter~~, thus the negative temperature offsets are not physically consistent. Snow depth for the two models is thick enough to produce a warming effect (Figure 6). This suggests problems with soil thermal conductivity that maintains deep soil warmth in those regions with a negative insulation effect~~. However, we observe that the snow depths from ISBA and JULES are not very thick (<10 cm) in most places on TP (Figure 6). Figure 9 shows the temperature offset between ground surface and air temperature as a function of snow depth. By inspection we note that there is different behavior for snow depths thinner and thicker than 4 cm. For snow depth > 4 cm, most negative offsets disappear in ISBA and JULES, which means that the ground surface temperature is warmer than air temperature for snow depth larger than 4 cm. For snow depth < 4 cm, the ground surface temperature of much of the region is colder than air temperature in ISBA and JULES, which indicates the cooling effect of thin snow. The very small or slightly negative temperature offset for thin snow is also seen in the other models. Of course, the strength of this effect depends on the individual model's simulation/parameterization of the snow processes (such as sublimation, evaporation, melting). The thin snow mechanism is also confirmed by the weak insulation effect in Figure 10. ~~Hence, although the permafrost maps produced by the models have a K > 0.2 compared with the observation-based Wang06 map, there are problems with the surface and soil temperature profiles.~~

6 Robustness of the results

6.1 Choice of thresholds in the methdologies

In Sect. 4 we used the most commonly applied threshold of each method, based on the empirical findings from previous studies, to compare models and methods. However, the thresholds themselves have the potential to affect the results. To reduce the latent uncertainties in terms of the methodologies, we also examine the sensitivity of permafrost area for different thresholds (Table 2), calculating changes in the permafrost area (Table 3) for a range of thresholds for each method (i.e., -3 ℃<MAAT<0 ℃; 0.4<F<0.6; 0.4<SFI<0.6; 0 ℃<MAGT<0.5 ℃).

Generally, when the permafrost definition requires colder climate, the derived permafrost area becomes smaller. The across-threshold uncertainty (Table 3) is similar for different models. But the across-threshold uncertainty with SFI varies greatly among models, $23 \sim 105 \times 10^4$ km^2, which is due to the seasonal amplitude of ground surface temperatures it requires. This is illustrated in Figure 5 where UVic and LPJ-GUESS have a relatively small seasonal amplitude of ground surface temperature, which corresponds to their small across-threshold variability for SFI derived area in Table 3.

The across-model uncertainty is highly consistent even with different thresholds for each method (Table 3 final column). Thus it seems changing the thresholds does not affect one key point in our paper: that across-model uncertainties using direct methods are much larger than using indirect ones. Large across-model uncertainties using direct methods imply that differences among these land surface processes are worthy of investigation.

6.2 Model settings

The lowest soil boundary is a critical uncertainty affecting the simulation of permafrost (Nicolsky et al., 2007). The common boundary of 3 m soil depth may produce uncertainties in the derived permafrost area. Three (CLM4.5, ISBA, UVic) of the six models extended the soil to deeper depths (Table 1), which provides insight on this issue. As UVic does not do a reasonable simulation of snow cover and ground temperature, we feel it is not necessary to include this model in the discussion here. Based on results from CLM4.5 and ISBA, the permafrost area calculated from MAGT at 3 m and at 10 m only changes by 1×10^4 km^2. For results from CLM4.5, the areas calculated from MAGT at 20 m and 30 m do not change from the one calculated at 10 m. This is due to MAGT only considering annual mean soil temperature, not the seasonal cycle. This is consistent with the finding that the across-threshold uncertainty for MAGT-derived permafrost area is quite small (Table 3). However, the derived permafrost area with the TSL method improves when soil depth used for calculation is increased from 3 m to 5 m (Table 6). This sensitivity is because TSL requires information on the seasonal cycle of soil temperature. In other words, results of TSL method are sensitive to the active layer dynamics. The permafrost on TP is usually much warmer and has a deeper active layer than found in continuous permafrost of the arctic and boreal region. Hence deeper soil layers would be well suited for TP permafrost simulation. A shallow column in a permafrost model can cause problems in the simulation of the degradation of warm permafrost (near 0^{o} C), which is expected for projections of future climate warming (Lawrence et al., 2008). In addition, Alexeev et al. (2007) pointed out that deep soil configuration can improve the simulation of seasonal and even annual cycle of shallow layers. Nicolsky et al. (2007) recommend a soil column of at least 80 m for models applied to permafrost regions.

Soil layer discretization and spatial resolutions are different among the six models (Table 1). In this study we linearly interpolated and extrapolated the soil temperatures onto the standard layers (Sect. 3.1). The impact of ground surface temperature extrapolation was found to be small by comparing Figures 7 and 8 with those made using temperatures at 5 cm depth (not shown), with both geographical patterns and widespread negative surface temperature offsets in ISBA and JULES. We re-gridded the Wang06 map onto each model's spatial resolution to evaluate the models objectively. This leads to an error bar estimate of half a grid cell area, up to 20 $\times 10^4$ km^2, which is half of the spread of observation area estimates (Sect. 3.2). Daily and hourly temperature data may make some differences to the permafrost extent map, but the diurnal cycle wave decays at shallower soil depths than the deepest model layer.

~~6~~7 Summary and Conclusions

Results of this model intercomparison quantify, for the first time, the uncertainties of model derived permafrost area on the Tibetan Plateau (TP). The uncertainties stem from across-model and across-diagnostic method variability as well as historic climate data uncertainties. According to the agreement of the air temperature based diagnostic methods (MAAT and F), we found lower uncertainty in permafrost area associated with air temperature forcing (99 to 135 $\times 10^4$ km^2) in comparison with the uncertainty (1 to 128 $\times 10^4$ km^2) associated with the simulation of soil temperature used in the other three diagnostic methods (TSL, MAGT, and SFI).~~;~~ The observation-based Wang06 permafrost area is 101 $\times 10^4$ km^2.

~~The~~ Most models in this study ~~generally~~ produced permafrost maps in better agreement with the Wang06 map using the MAGT and SFI methods rather than with the TSL method. But this does not mean that the models simulate permafrost dynamics correctly. Although most models can capture the threshold value of MAGT and SFI, their ground temperatures still show various biases, both in the mean annual value and the seasonal variation. Therefore, most models produce worse permafrost maps with the TSL method. The TSL method is a more demanding, and to date, elusive target.

Modeled snow depth and surface and soil temperature offsets vary widely amongst the models. If the observation sites for soil temperature are representative, then LPJ-GUESS and UVic have substantial biases in their soil temperature simulations, mainly attributable to inappropriate description of the surface (vegetation, snow cover) and soil properties (soil texture, hydrology). Other models (ISBA, JULES) show biases in the simulation of winter soil temperature.

~~From investigations in the arctic and boreal regions, we know that the specification of surface and soil properties needs substantial improvement. In addition, models need to consider deeper soil columns in their simulations. Nicolsky et al. (2007) recommend a soil column of at least 80 m for models applied to arctic and boreal regions. The permafrost in the TP is usually much warmer and with a deeper active layer than found in continuous permafrost of the arctic and boreal region, hence deep soil layers would also be applicable for TP permafrost simulation. A shallow column in a permafrost model can cause problems in the simulation of the degradation of warm permafrost (near 0^{o} C), which is expected for projections of future climate warming (Alexeev et al., 2007; Lawrence et al., 2008).~~

Further evaluation of model results from the permafrost-RCN is underway for TP that examines permafrost temperature, active layer thickness and carbon balance under present and future climate forcing. We also plan to complement this model intercomparison study by an uncertainty quantification analysis of key model parameters (e.g. improved vegetation and snow albedo, soil colors, etc) with the CoLM model. However, a crucial requirement for this is much better data availability allowing for better spatial coverage across the TP in the evaluation of simulated ground temperature profiles. Under the Chinese Scientific Foundation Project “Permafrost Background Investigation on the Tibetan Plateau” (No. 2010CB951402), a series of new stations have been established, especially in the depopulated zone. More ground truth data will be published in the near future, which will also be assimilated in a new observation-based permafrost map.

Acknowledgements

This study was supported by the Permafrost Carbon Vulnerability Research Coordination Network, which is funded by the National Science Foundation. 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 Programme (GA01101) and the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n°282700. This research was also sponsored by Chinese foundations: (1) the National Basic Research Program of China (Grant No. 2015CB953600), (2) the National Science Foundation of China (Grant No. 40905047), (3) the National Natural Science Foundation of China (Grant No.41275003), and (4) the National Natural Science Foundation of China (Grant No.41030106). In addition, Bertrand Decharme and Christine Delire were supported by the French Agence Nationale de la Recherche under agreement ANR-10-CEPL-012-03.

References

Alexeev, V., Nicolsky, D., Romanovsky, V., and Lawrence, D.: An evaluation of deep soil configurations in the CLM3 for improved representation of permafrost, *Geophys. Res. Lett.*, **34**, L09502, doi:10.1029/2007GL029536, 2007.

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.

AWFA: Data format handbook for AGRMET, Available online: http://www.mmm.ucar. edu/mm5/documents/DATA_FORMAT_HANDBOOK.pdf (accessed on January 20, 2010), 2002.

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.

Brown, J., Ferrians, O., Heginbottom, J., and Melnikov, E.: Circum-Arctic map of permafrost and ground-ice conditions, US Geological Survey Reston, 1997.

Che, T., Li, X., Jin, R.: Armstrong R, Zhang TJ, 2008. Snow depth derived from passive microwave remote-sensing data in China, Annals of Glaciology, 49:145-154, 2008.

Cohen, J.: A Coefficient of Agreement for Nominal Scales, Educational and Psychological Measurement, 20 (1), 37-46, 1960.

Cheng, G., and Jin, H.: Permafrost and groundwater on the Qinghai-Tibet Plateau and in northeast China[J], Hydrogeology Journal, 21(1): 5-23, 2013.

Dai, Y., et al.: 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., Boone, A., Delire, C., and Noilhan, J.: Local evaluation of the Interaction between Soil Biosphere Atmosphere soil multilayer diffusion scheme using four pedo-transfer functions, J. Geophys. Res.-Atmos., 116, D20126,doi:10.1029/2011JD016002, 2011.

Dörfer, C., Kühn, P., Baumann, F., He, J., and Scholten, T.: Soil organic carbon pools and stocks in permafrost-affected soils on the Tibetan Plateau[J], PloS one, 8(2): e57024, doi:10.1371/journal.pone.0057024, 2013.

Essery, R., Morin, S., Lejeune, Y., and Ménard, C. B.: A comparison of1701snow models using observations from an alpine site, Adv. WaterResour., 55, 131–148, doi:10.1016/j.advwatres.2012.07.013, 2013.

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

Gruber, S.: Derivation and analysis of a high-resolution estimate of global permafrost zonation, The Cryosphere, Volume 6, Issue 1, pp. 221-233, 6, 221-233, 2012.

Gubler, S., Endrizzi, S., Gruber, S., and Pursves, R. S..: Sensitivities and uncertainties of modeled ground temperatures in mountain environments[J], Geoscientific Model Development Discussions, 6(1): 791-840, 2013.

Guo, D., and Yang, M.: Simulation of Soil Temperature and Moisture in Seasonally Frozen Ground of Central Tibetan Plateau by SHAW Model, Plateau Meteorology, 29 (6), 1369-1377, 2010.

Guo, D., Wang, H., and Li, D.: A projection of permafrost degradation on the Tibetan Plateau during the 21st century, Journal of Geophysical Research: Atmospheres (1984–2012), 117, D05106, doi:10.1029/2011JD016545, 2012.

Guo, D., and Wang, H.: Simulation of permafrost and seasonally frozen ground conditions on the Tibetan Plateau, 1981–2010, Journal of Geophysical Research: Atmospheres, 118, 5216-5230, 2013.

Heginbottom, J.: Permafrost mapping: a review, Progress in Physical Geography, 26, 623-642, 2002.

Hillel, D.: Environmental soil physics: Fundamentals, applications, and environmental considerations [M], Academic Press, New York, USA, 1998.

IPCC 2013 Climate change 2013: The Physical Science Basis Contribution of Workin g Group I to the Fifth Assessment Report of the Intergovernmental Panel on Cli mate Change, Cambridge University Press, Cambridge.

Ji, D., Wang, L., Feng, J., Wu, Q., Cheng, H., Zhang, Q., Yang, J., Dong, W., Dai, Y., Gong, D., Zhang, R.-H., Wang, X., Liu, J., Moore, J. C., Chen, D., and Zhou, M.: Description and basic evaluation of Beijing Normal University Earth System Model (BNU-ESM) version 1, Geosci. Model Dev., 7, 2039-2064, doi:10.5194/gmd-7-2039-2014, 2014.

Jin, H., Yu, Q., Lü, L., Guo, D., He, R., Yu, S., Sun, G., and Li, Y.: Degradation of permafrost in the Xing'anling Mountains, Northeastern China, Permafrost and Periglacial Processes, 18, 245-258, 2007.

Koven, C., Riley, W., and Stern, A.: Analysis of permafrost thermal dynamics and response to climate change in the CMIP5 Earth System Models, Journal of Climate, 26, 1877-1900, 2013.

Lawrence, D., and Slater, A.: A projection of severe near‐surface permafrost degradation during the 21st century, Geophysical Research Letters, 32, L24401, doi:10.1029/2005GL025080, 2005.

Lawrence, D., Slater A., Romanovsky V., and Nicolsky D.: The sensitivity of a model projection of near-surface permafrost degradation to soil column depth and inclusion of soil organic matter, J. Geophys. Res., 113, F02011, doi:10.1029/2007JF000883, 2008.

Li, Q., Sun, S., and Dai, Q.: The numerical scheme development of a simplified frozen soil model, Advances in Atmospheric Sciences, 26, 940-950, 2009.

Li, S., and Cheng, G.: Map of Frozen Ground on Qinghai -Xizang Plateau, Lanzhou: Gansu Culture Press, 1996.

Li, S., Gao, S., Yang, P., and Chen, H.: Some problems of freeze-thaw desertification in the Qinghai-Tibetan Plateau: a case study on the desertification regions in the western and northern Tibet[J], J Glaciol Geocryol, 27(4): 476-485, 2005.

Li, Z., Tang, P., Zhou, J., Tian, B., Chen, Q., and Fu, S.: Permafrost environment monitoring on the Qinghai-Tibet Plateau using time series ASAR images[J], International Journal of Digital Earth, DOI:10.1080/17538947.2014.923943, 2014.

Luo, S., Lv, S., Zhang, Y., Hu, Z., Ma, Y., Li, S., and Shang, L.: Simulation analysis on land surface process of BJ site of central Tibetan Plateau Using CoLM, Plateau Meteorology, 27, 259-271, 2008.

McGuire, D., et al.: An retrospective assessment of the vulnerability of permafrost carbon in the earth system: Comparison of dynamics among process-based models, in preparation, 2014.

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.

Monserud, R., and Leemans, R.: Comparing global vegetation maps with the Kappa statistic, Ecological Modelling, 62, 275-293, 1992.

Nan, Z., Huang, P., and Zhao, L.: Permafrost distribution modeling and depth estimation in the Western Qinghai-Tibet Plateau, Acta Geographica Sinica, 68, 318-327, 2013.

Nan, Z., Li, S., and Liu, Y.: Mean annual ground temperature distribution on the Tibetan Plateau: permafrost distribution mapping and further application, Journal of Glaciology and Geocryology, 24, 142-148, 2002.

Nan, Z., Li, S., Cheng, G., and Huang, P.: Surface frost number model and its application to the Tibetan plateau, Journal of Glaciology and Geocryology, 34, 89-95, 2012.

Nelson, F., and Outcalt, S.: A computational method for prediction and regionalization of permafrost, Arctic and Alpine Research, 19(3), 279-288, 1987.

Nicolsky, D. J., V. E. Romanovsky, V. A. Alexeev, and D. M. Lawrence: Improved modeling 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.

Peng, S., Ciais, P., Krinner, G., Wang T., Gouttevin I., McGuire A.D., Lawrence D., Burke E., Chen X., Delire C., Koven C., MacDougall A., Rinke A., Saito K., Zhang W., Alkama R., Bohn T.J., Decharme B., 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 Discuss., 9, 2301–2337, doi:10.5194/tcd-9-2301-2015, 2015.

Qin, J., Yang, K., Liang, S., Zhang, H., Ma, Y., Guo, X., and Chen, Z.: Evaluation of surface albedo from GEWEX‐SRB and ISCCP‐FD data against validated MODIS product over the Tibetan Plateau, Journal of Geophysical Research: Atmospheres (1984–2012), 116, D24116, doi:10.1029/2011JD015823, 2011.

Ran, Y., Li, X., Cheng, G., Zhang, T., Wu, Q., Jin, H., and Jin, R.: Distribution of permafrost in China: An overview of existing permafrost maps, Permafrost and Periglacial Processes, 23, 322-333, 2012.

Rawlins, M.A., McGuire, A.D., Kimball, J.S., 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 CO2 Exchange Across Northern Eurasia, Biogeosciences, 12, 4385-4405, doi:10.5194/bg-12-4385-2015, 2015.

Saito, K., Sueyoshi, T., Marchenko, S., Romanovsky, V., Otto-Bliesner, B., Walsh, J., Bigelow, N., Hendricks, A., and Yoshikawa, K.: LGM permafrost distribution: how well can the latest PMIP multi-model ensembles perform reconstruction?, Climate of the Past, 9, 1697-1714, 10.5194/cp-9-1697-2013, 2013.

Schuur, E. A., Bockheim, J., Canadell, J. G., Euskirchen, E., Field, C. B., Goryachkin, S. V., Hagemann, S., Kuhry, P., Lafleur, P. M., and Lee, H.: Vulnerability of permafrost carbon to climate change: Implications for the global carbon cycle, BioScience, 58, 701-714, 2008.

Shi, Y., Mi, D., Feng, Q., Li, P., and Wang, Z.: Map of snow, ice and frozen ground in China, China Cartographic Publishing House, Beijing, China, 1988.

Slater, A., and Lawrence, D.: Diagnosing Present and Future Permafrost from Climate Models, Journal of Climate, 26, 5608–5623, doi: http://dx.doi.org/10.1175/JCLI-D-12-00341.1,2013.

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.

Tian, L., Li, W., Zhang, R., Tian, L., Zhu, Q., Peng, C., and Chen, H.: The Analysis of Snow Information from 1979 to 2007 in Qinghai-Tibetan Plateau, Acta Ecologica Sinica, 34 (20) : 5974-5983, 2014.

Van Duin, R. H. A.: The influence of management on the temperature wave near the surface, Tech. Bull. 29, 21 pp., Inst. of Land and Water Manage. Res.,Wageningen, Netherlands, 1963.

Wang, C., and Shi, R.: Simulation of the land surface processes in the Western Tibetan Plateau in summer, Journal of Glaciology and Geocryology, 29 (1), 73-81, 2007.

Wang, G., Li, Y., Wang, Y., and Wu, Q.: Effects of permafrost thawing on vegetation and soil carbon pool losses on the Qinghai–Tibet Plateau, China, Geoderma, 143, 143-152, 2008.

Wang, K., Cheng, G., Jiang, C., and Niu, F.: Variation of Thermal Diffusivity and Temperature Simulation of Soils of Vertical Heterogeneity in Nagqu Prefecture in the Tibetan Plateau [J], Journal of Glaciology and Geocryology, 29 (3), 470-474, 2007.

Wang, T., Wang, N. L., and Li, S. X.: Map of the glaciers, frozen ground and desert in China, 1: 4,000,000 (in Chinese) [J], Chinese Map Press, Beijing, China, 2006.

Wang, X., Yang, M., and Wan, G.: Processes of Soil Thawing-Freezing and Features of Ground Temperature and Moisture at D105 on the Northern Tibetan Plateau, Journal of Glaciology and Geocryology, 34, 56-63, 2012.

Wang, Z.: Applications of permafrost distribution models on the Qinghai-Tibetan Plateau, Lanzhou University, 2010.

Wang, Z., Wang, X., and Li, Y.: Analyses of snow cover based on passive microwave remote sensing data and observed data over the Tibetan Plateau [J], Journal of Glaciology and Geocryology, 35(4), 783-792, 2013.

Wu, Q., and Liu, Y.: Ground temperature monitoring and its recent change in Qinghai–Tibet Plateau, Cold Regions Science and Technology, 38, 85-92, 2004.

Wu, Q., and Niu, F.: Permafrost changes and engineering stability in Qinghai-Xizang Plateau[J], Chinese Science Bulletin, 58(10): 1079-1094, 2013.

Wu, Q., and Zhang, T.: Changes in active layer thickness over the Qinghai - Tibetan Plateau from 1995 to 2007, Journal of Geophysical Research: Atmospheres (1984–2012), 115, D09107, doi:10.1029/2009JD012974, 2010.

Wu, T., Zhao, L., Li, R., Wang, Q., Xie, C., and Pang, Q.: Recent ground surface warming and its effects on permafrost on the central Qinghai - Tibet Plateau[J], International Journal of Climatology, 33(4): 920-930, 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.

Xin, Y., Wu, B., Bian, L., Liu, G., Zhang, L., and Li, R.: Simulation study of permafrost hydro-thermo dynamics on Asian climate (in Chinese), the 29th annual meeting of Chinese Meterological Society, Shenyang, China, 2012-9-12, P461, 2012.

Xiong, J., Zhang, Y., Wang, S., SHang, L., CHen, Y., and SHen, X.: 1. Key Laboratory of Land Surface Process and Climate Change in Cold and Arid Regions, Cold and Arid Regions Environment Research Institute, Chinese Academy of Science, Lanzhou 730000, China; 2. University of Chinese Academy of Sciences, Beijing 100049, China, PLATEAU METEOROLOGY, 33, 323-336, 2014.

Xu, X., and Lin, Z.: Remote sensing retrieval of surface monthly mean albedo in Qinghai - Xizang Plateau, Plateau Meteorology, 21, 233-237, 2002.

Yao, T., Qin, D., Shen, Y., Zhao, L., Wang, N., and Lu, A.: Cryospheric changes and their impacts on regional water cycle and ecological conditions in the Qinghai Tibetan Plateau[J], Chinese J. Nature, 35: 179-186, 2013.

Yang, M., Nelson, F., Shiklomanov, N., Guo, D., and Wan, G.: Permafrost degradation and its environmental effects on the Tibetan Plateau: A review of recent research, Earth-Science Reviews, 103, 31-44, 2010.

Yang, M., Yao, T., Gou, X., Koike, T., and He, Y.: The soil moisture distribution, thawing–freezing processes and their effects on the seasonal transition on the Qinghai–Xizang (Tibetan) plateau, Journal of Asian Earth Sciences, 21, 457-465, http://dx.doi.org/10.1016/S1367-9120(02)00069-X, 2003.

Yang, M., Yao, T., and Gou, X.: The freezing thawing processes and hydro-thermal characteristics along the road on the Tbietan Plateau, Advancement of Natural Scinece (China), 10 (5), 443-450, 2000.

Yang, M., Yao, T., and He, Y.: The extreme value analysis of the ground temperature in northern part of Tibetan Plateau records from D110 site, Journal of Mountain Science (China), 17 (3), 207-211, 1999.

Yu, F., Qi, J., Yao, X., and Liu, Y.: In-situ monitoring of settlement at different layers under embankments in permafrost regions on the Qinghai–Tibet Plateau[J], Engineering Geology, 160: 44-53, 2013.

Zhang, W., Wang, G., Zhou, J., Liu, G., and Wang, Y.: Simulating the Water-Heat Processes in Permafrost Regions in the Tibetan Plateau Based on CoupModel, Journal of Glaciology and Geocryology, 34, 1099-1109, 2012.

Zhang, Z., and Wu, Q.: Predicting changes of active layer thickness on the Qinghai-Tibet Plateau as climate warming, Journal of Glaciology and Geocryology, 34 (3), 505-511, 2012.

Zhou, Y., Guo, D., Qiu, G., Cheng, G., and Li, S.: China Permafrost, Science Press, Beijing. p. 232. (inChinese), 2000.

Tables

Table 1. The six land surface models, analyzed over the Tibetan plateau (TP)

Model	Native Resolution	Number of soil layers	Depth of soil column (m)	Spatial domain	Atmospheric Forcing Data
CLM4.5 Swenson and Lawrence, 2012 Oleson et al., 2013	1°×1.25°	30	38.1	Whole TP	CRUNCEP4[1]
CoLM Dai et al., 2003 Ji et al., 2014	1°×1°	10	2.86	Whole TP	Princeton[2]
ISBA Decharme et al. 2011	0.5°×0.5°	14	10	Permafrost region follow IPA map	WATCH [3]
JULES Best et al., 2011	0.5°×0.5°	30	2.95	Whole TP	WATCH [3]
LPJ-GUESS Gerten et al., 2004 Wania et al., 2009	0.5°×0.5°	25	3	Permafrost region follow IPA map	CRU TS 3.1[4]
UVic Meissner et al., 2003	1.8°×3.6°	14	198.1	Whole TP	CRUNCEP4[1]

[1]Viovy and Ciais (http://dods.extra.cea.fr/)

[2]Sheffield et al. (2006) (http://hydrology.princeton.edu/data.pgf.php)

[3]Weedon et al. (2011) (http://www.waterandclimatechange.eu/about/watch-forcing-data-20th-century)

[4]Harris et al. (2013), University of East Anglia Climate Research Unit (2013)

~~Table 2.The five diagnostic methods to derive permafrost~~

~~Method~~	~~Definition~~	~~Data used for calculation~~
~~TSL~~	~~More than 24 consecutive months soil temperature ≤ 0℃~~	~~0 ~ 3m monthly soil temperature~~
~~MAGT~~	~~Mean annual of 3 m soil temperature ≤ 0℃~~	~~Mean annual of 3 m soil temperature~~
~~SFI~~	~~Surface frost number ≥ 0.5~~	~~Annually maximum and minimum ground surface temperature~~
~~F~~	~~Air frost number ≥ 0.5~~	~~Annually maximum and minimum air temperature~~
~~MAAT~~	~~Mean annual air temperature ≤ -2℃~~	~~Mean annual of air temperature~~

Table 2. The five diagnostic methods and threshold values used to derive permafrost. The thresholds commonly used in the literature and in this paper are marked in bold.

Method	Definition	Threshold	Data used for calculation
TSL	More than 24 consecutive months soil temperature ≤ a threshold	**0 ℃**	0 ~ 3m monthly soil temperature
MAGT	Mean annual of 3 m soil temperature ≤ a threshold	**0 ℃**, 0.5 ℃	Mean annual of 3 m soil temperature
SFI	Surface frost number ≥ a threshold	0.4, **0.5**, 0.6	Annually maximum and minimum ground surface temperature
F	Air frost number ≥ a threshold	0.4, **0.5**, 0.6	Annually maximum and minimum air temperature
MAAT	Mean annual air temperature ≤ a threshold	0 ℃, -1 ℃, **-2 ℃**, -3 ℃	Mean annual of air temperature

~~Table 3. Derived permafrost area inside the common modeling region on Tibetan plateau (10^4 km^2) from 6 LSMs and 5 diagnostic methods~~

		~~CLM4.5~~	~~CoLM~~	~~JULES~~	~~UVic~~	~~ISBA~~	~~LPJ-GUESS~~	~~across-model uncertainty~~
~~Indirect method~~	~~MAAT~~	~~113~~	~~105~~	~~111~~	~~99~~	~~109~~	~~110~~	~~14~~
	~~F~~	~~135~~	~~127~~	~~131~~	~~118~~	~~130~~	~~131~~	~~17~~
~~Direct method~~	~~TSL~~	~~60~~	~~1~~	~~62~~	~~8~~	~~44~~	~~119~~	~~118~~
	~~MAGT~~	~~104~~	~~88~~	~~96~~	~~8~~	~~61~~	~~128~~	~~120~~
	~~SFI~~	~~115~~	~~62~~	~~100~~	~~8~~	~~112~~	~~119~~	~~111~~
~~across-direct method uncertainty~~		~~55~~	~~87~~	~~38~~	~~0~~	~~68~~	~~9~~	

Table 3. Derived permafrost area inside the common modeling region on Tibetan plateau (10^4 km^2) from 6 LSMs and 5 diagnostic methods, using different thresholds. The results of thresholds commonly used in the literature and in this paper are marked in bold.

		CLM4.5	CoLM	JULES	UVic	ISBA	LPJ-GUESS	across-model uncertainty
Indirect method	MAAT≤ 0 ℃	130	124	126	116	127	129	14
	MAAT≤ -1 ℃	122	117	119	109	119	120	13
	MAAT≤ -2 ℃	**113**	**105**	**111**	**99**	**109**	**110**	**14**
	MAAT≤ -3 ℃	95	83	96	81	91	93	15
	across-threshold uncertainty	35	41	30	35	36	36	
	F≥ 0.4	140	135	138	126	138	138	14
	F≥ 0.5	**135**	**127**	**131**	**118**	**130**	**131**	**17**
	F≥ 0.6	117	93	106	89	100	101	28
	across-threshold uncertainty	23	42	32	37	38	37	
Direct method	**TSL**	**60**	**1**	**62**	**8**	**44**	**119**	**118**
	MAGT≤ 0.5 ℃	112	102	104	8	72	131	123
	MAGT≤ 0 ℃	**104**	**89**	**96**	**8**	**61**	**128**	**120**
	across-threshold uncertainty	8	13	8	0	11	3	
	SFI≥ 0.4	135	122	130	32	131	127	103
	SFI≥ 0.5	**116**	**62**	**100**	**8**	**113**	**119**	**111**
	SFI≥ 0.6	42	17	38	4	55	104	100
	across-threshold uncertainty	93	105	92	28	76	23	
across-direct method uncertainty (based on commonly used methods TSL, MAGT≤0℃, SFI≥0.5)		**56**	**88**	**38**	**0**	**69**	**9**	

Table 4. Model - observed temperatures differences in mean annual and seasonal cycle amplitude of air and soil temperature, based on data from 1996-2000 (section 3.4; Figure 4), and the corresponding vegetation and soil properties of both observation and models. Air temperature data is only available for D66 station and limited from 1997/9 to 1998/8. Thus the statistics of ground temperature of D66 is also confined to this period .

D66 (35.63 °N, 93.81 °E)									
	Temperature bias "Model - Observation"						Soil conditions		
	Air temperature		Ground temperature						
			At 0.04 m depth		At 2.63 m depth		Bare ground	Vegetation	Texture (top soil)
	Mean annual	Seasonal amplitude	Mean annual	Seasonal amplitude	Mean annual	Seasonal amplitude			
Obs[1]							100%	None	gravel
CLM4.5[2]	4.3	1	2	-0.2	2	3.5	81%	10% boreal shrub 8% C3 arctic grass	63% sand 19% clay
CoLM[3]	2.3	0.1	0	0.1	-1	2.4	87%	4% boreal shrub 5% C3 arctic grass 3% C3 non arctic grass	43% sand 18% clay
ISBA[4]	1.4	0.1	-1.3	-1.3	0.8	0.5	53%	46% C3 grass	55% sand 7% clay
JULES[#]	1.1	0.3	-0.5	2.1	-2	4			
LPJ -GUESS[*5]	1.5	-0.1	-3.4	-6.6	-3.7	1.5		tundra	clay-like
UVic[6]	2.6	0.5	7.5	-1.5	7.6	2.1	100%	None	44% sand 24% clay

D105 (33.07 °N, 91.94 °E)					
	Temperature bias “Model - Observation”		Soil conditions		
	Ground temperature				
	At 3 m depth		Bare ground	Vegetation	Texture (top soil)
	Mean annual	Seasonal amplitude			
Obs[7]			50%-60%	grass (Leontopodium nanum)	coarse and fine sand
CLM4.5[2]	-1.2	0.8	48%	17% boreal_shrub 30% C3 arctic grass	60% sand 20% clay
CoLM[3]	0.1	0.2	7%	69% C3 arctic grass 24% C3 non arctic grass	38% sand 16% clay
ISBA[4]	0.9	-0.9	27%	72% C3 grass	52% sand 10% clay
JULES[#]	-1.8	1.8			
LPJ -GUESS[*5]	-3.7	0.7		tundra	clay-like
UVic[6]	1	-0.2	7%	33% C3 grass 60% shrub	43% sand 32% clay

D110 (32.82 °N, 93.01 °E)					
	Temperature bias "Model - Observation"		Soil conditions		
	Ground temperature				
	At 0.04 m depth				
	Mean annual	Seasonal amplitude	Bare ground	Vegetation	Texture (top soil)
Obs[8]			60-70%	grass (Kobresia humilis)	coarse and fine sand
CLM4.5[2]	-1.8	1	33%	7% boreal_shrub 57% C3 arctic grass	60% sand 21% clay
CoLM[3]	0.5	1.4	1%	56% C3 arctic grass 43% C3 non arctic grass	45% sand 17% clay
ISBA[4]	-1.4	0.8	10%	89% C3 grass	50% sand 11% clay
JULES[#]	-1.9	0.9			
LPJ -GUESS[*5]	-4.1	-3.7		tundra	clay-like
UVic[6]	1.1	-0.5	6%	31% C3 grass 60% shrub	45% sand 30% clay

[1]Yang et al. (2000)

[2]https://dl.dropboxusercontent.com/u/41730762/surfdata_0.9x1.25_simyr1850_c130415.nc

[3] Dai et al. (2003); Ji et al. (2014)

[4]Harmonized World Soil Database

[5]Thermal diffusivities follow Van Duin (1963) and Jury et al. (1991), volumetric fraction of organic material follow Hillel (1998), water held below wilting point and porosity from AWFA (2002)

[6]Scholes and de Colstoun (2012) (http://www.daac.ornl.gov)

[7] Wang et al. (2012)

[8]Yang et al. (1999)

[*] The classification of soil texture is based on soil volumetric water holding capacity, thermal diffusivities, volumetric fraction of organic material, water held below wilting point and porosity

[#]This model doesn't provide soil parameter information

Table 5. Description of Model Characteristics Relevant to Soil Temperatures ~~Year-round relative Relative model Model characteristics Characteristics~~ on TP

Model	Snow cover[1]	Albedo[2]	Soil water[3]	Unfrozen water effect during phase change[4]	Surface Organic layer insulation	Snow scheme[5]
CLM4.5	Medium	Medium	Medium	Yes	Yes	Dynamic & ML
CoLM	Medium	Medium	Medium	No	No	Dynamic & ML
ISBA	Low	Low	Medium	Yes	Yes	Static &ML
JULES	Low	Low	Medium	Yes	No	Dynamic & ML
LPJ-GUESS	Medium	Low	High	No	No	Static & BL
UVic	None	Low	High	Yes	No	Static & I

[1] Low snow cover is confined to high elevations, medium tends to be on western TP

[2] LPJ-GUESS has constant albedo everywhere and UVic albedo varies slightly due to vegetation, year-round albedo variability for other models depends mainly on snow cover in winter and soil moisture, vegetation, etc in summer

[3] soil water content includes both liquid and ice fractions

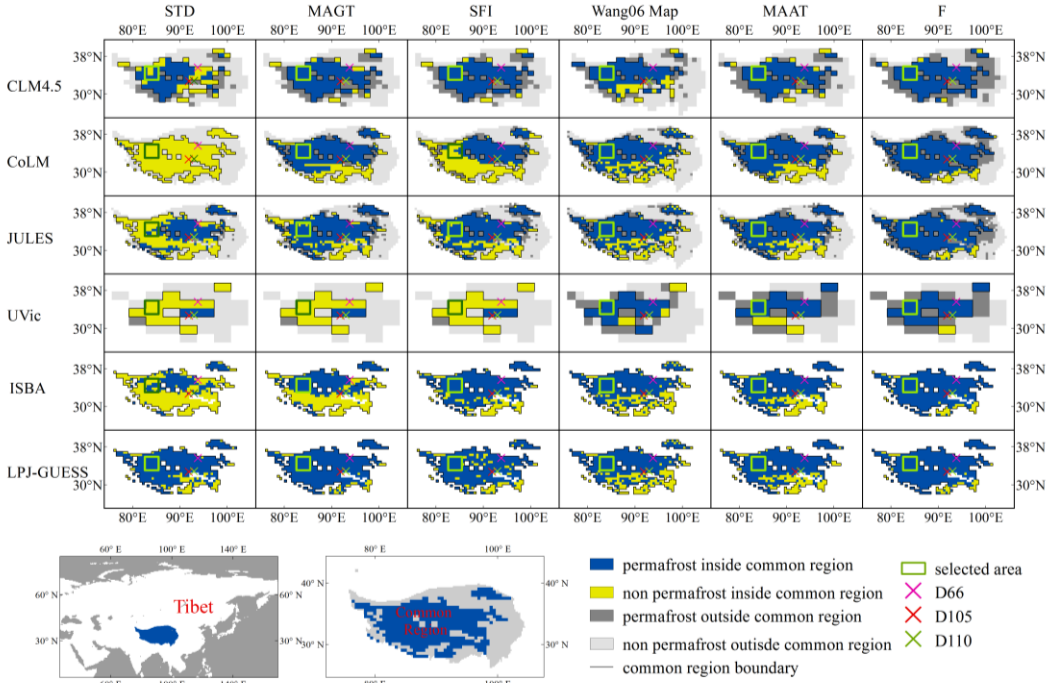
[4] all models calculate soil thermal properties depending on soil moisture and also phase change of water, but CoLM and LPJ-GUESS ignore solute dependent freezing processes

[5] Dynamic or static snow layering; ML: Multi-layer, BL: Bulk-layer, I: Implicit; according to *Slater et al*. [2001]

Table 6. Derived permafrost area (10^4 km^2) with deeper soil layers using the TSL method. The results for thresholds commonly used in the literature and in this paper are marked in bold.

Depth of deepest layer used for calculation	CLM4.5	ISBA
3 m	**60**	**44**
5 m	85	54

Figure Captions



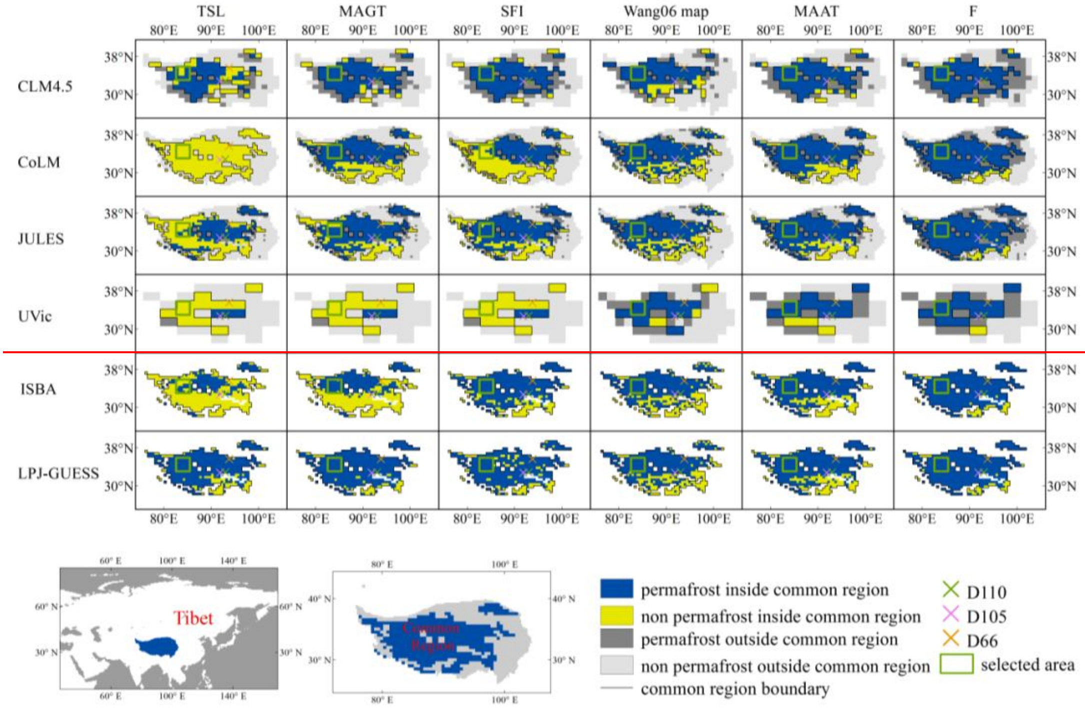
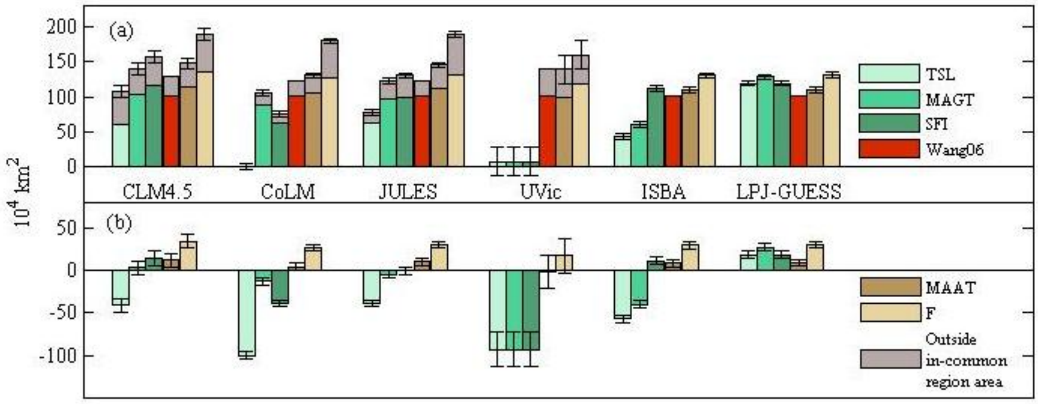


Figure 1. Permafrost maps derived from different diagnostic methods and models compared with Wang06 map. Permafrost inside the common modeling region is used for all-models inter-comparison, while permafrost outside allows further evaluation over the whole TP for CLM4.5, CoLM, JULES and UVic. The observation-based map

of permafrost (Wang et al., 2006) is re-gridded to match model resolution. The selected area in the western TP (33 °- 36 °N, 82.5 °- 85.5 °E) is used to examine across-model differences in Figure 5. Insets show location map of TP and how the common region is related to the TP.



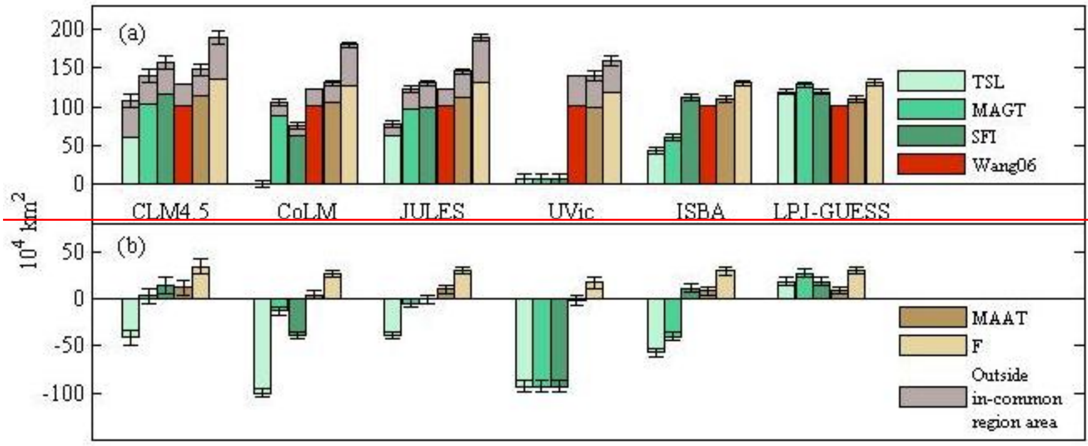


Figure 2. Permafrost areas derived from different diagnostic methods compared with Wang06 map. (a) Permafrost area, with TP permafrost outside the common region denoted by grey extensions to the bars for CLM4.5, CoLM, JULES and UVic. (b) Bias in permafrost area "Model minus Wang06 estimate", only for the common modeling region. The error bar is calculated as half of the averaged grid cell area of the model, so is model resolution dependent. Error bar is estimated from resolution differences.

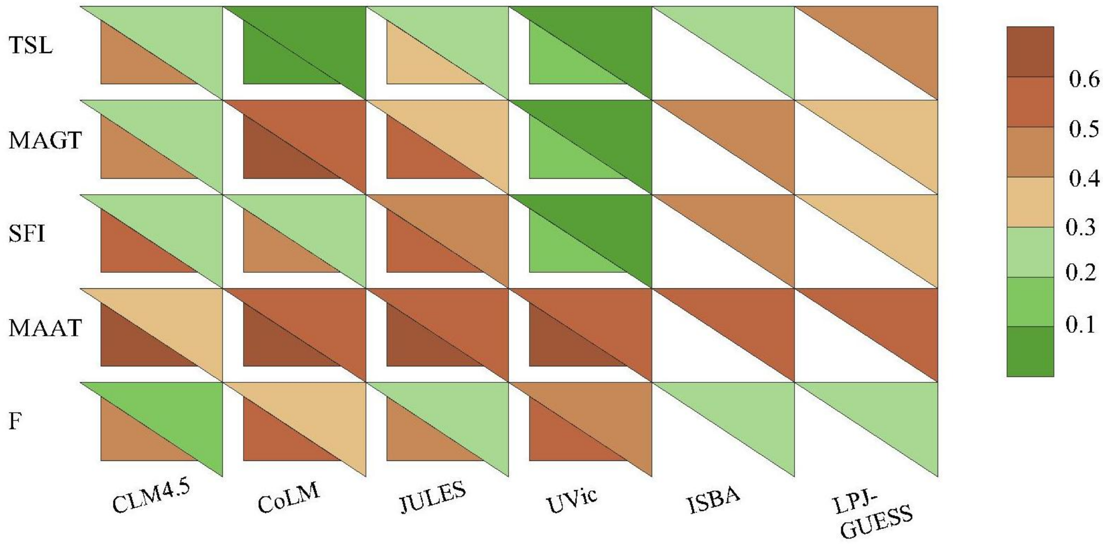
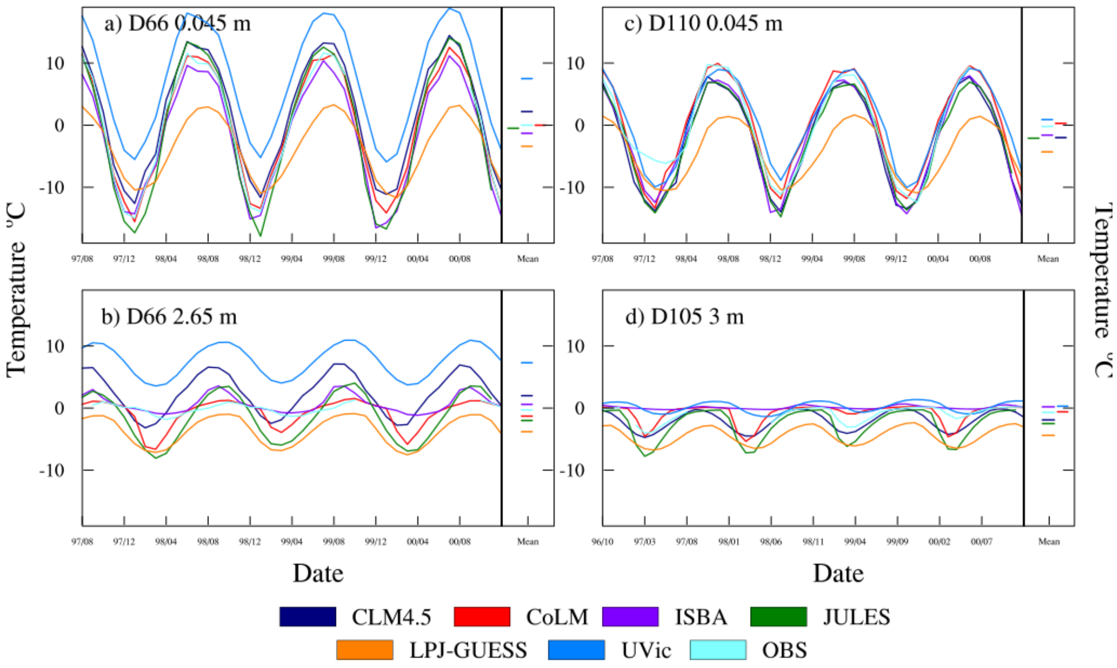


Figure 3. Kappa coefficient, *K*, quantifying the agreement between model-derived and Wang06 maps (see section 3.3). $K \geq 0.2$ indicates at least fair agreement with Wang06 map. The lower triangle is *K* for the whole TP and is only available for CLM4.5, CoLM, JULES and UVic, while the upper triangle is *K* for the common modeling region.



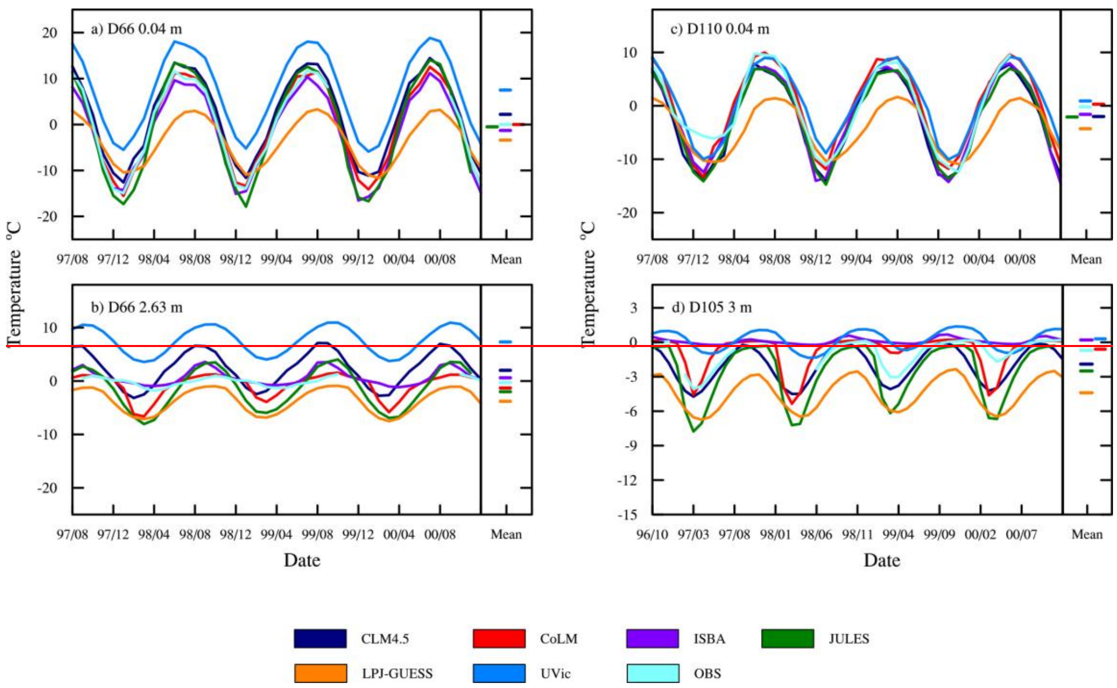
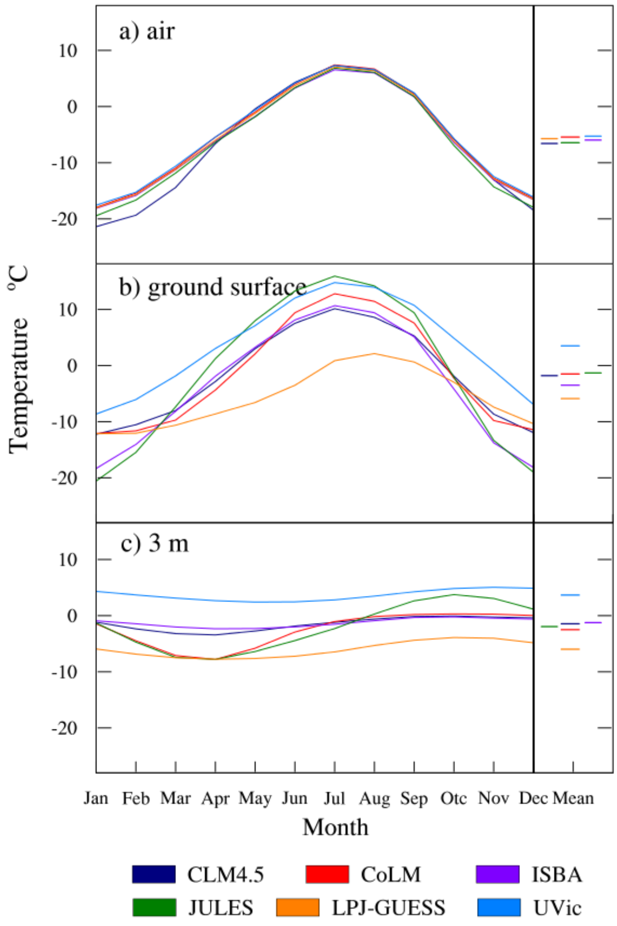


Figure 4. Monthly soil temperature variations at 3 stations from models and observations. (a) and (c) soil temperature of top layer. (b) and (d) soil temperature of deeper layer, 1996-2000. "Mean" denotes annual average temperature. We use the topmost available soil temperatures- (0.04 m at D66 and D110, no good data for D105) and lowest available ones (2.63 m at D66, 3 m of D105), while D110 has only temperatures at 2 m depth.


a) air
b) ground surface
c) 3 m
10
0
-10
-20
Temperature °C
Jan Feb Mar Apr May Jun Jul Aug Sep Otc Nov Dec Mean
Month
CLM4.5
CoLM
ISBA
JULES
LPJ-GUESS
UVic

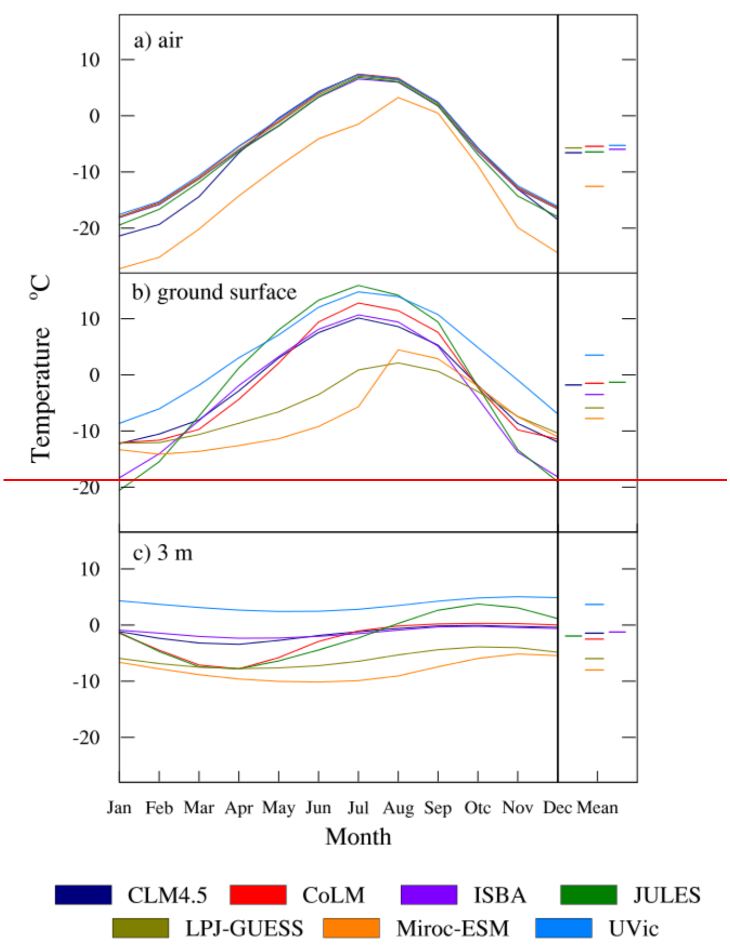
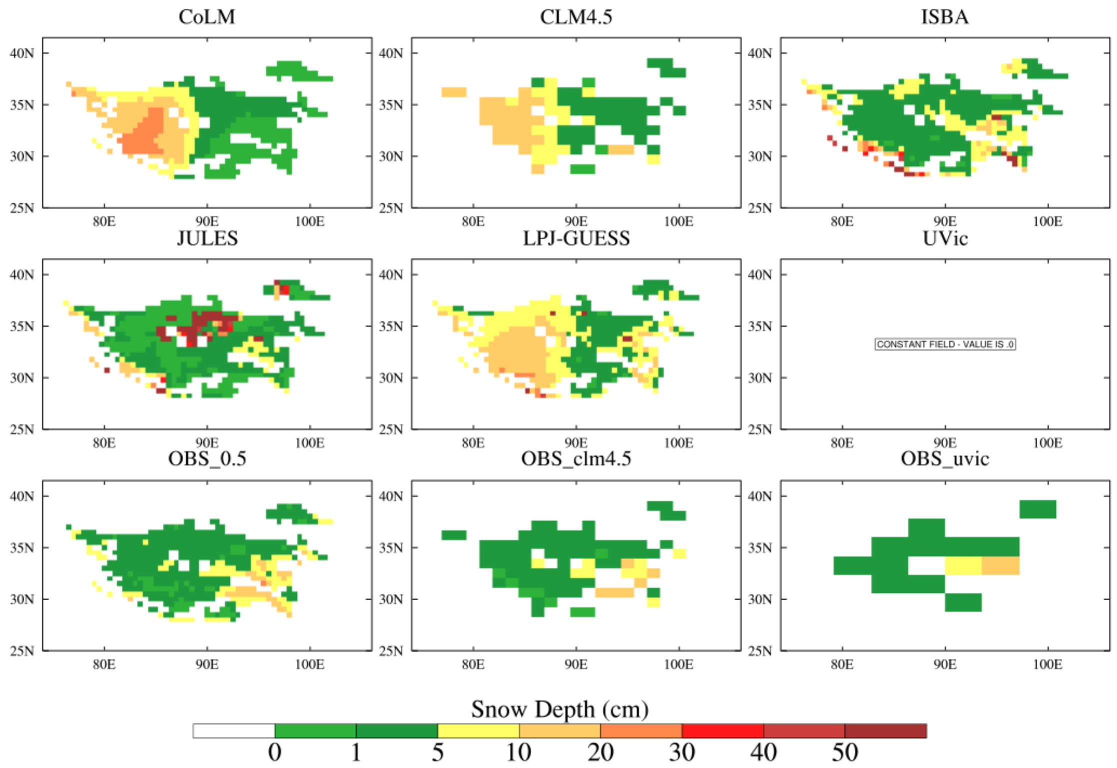


Figure 5. Monthly temperatures averaged over the selected western TP area in Figure 1. (a) Forcing air temperature, (b) Ground surface temperature, (c) 3 m soil temperature, averaged over 1980-2000.“ Mean” denotes annual average temperature.



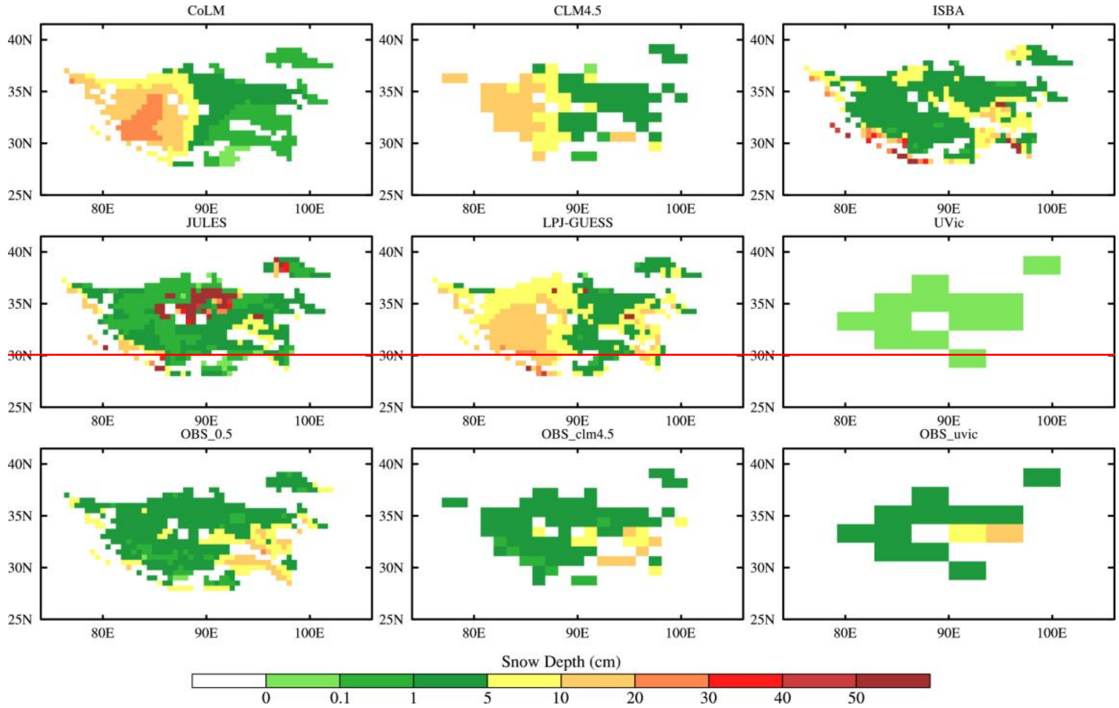


Figure 6. Winter snow depth for the common region, averaged over 1980-2000. Note the nonlinear color scale. We use the Long Time Series Snow Dataset of China (Che et al., 2008) (http://westdc.westgis.ac.cn) as observed snow depth. The observed snow depth plot is further interpolated onto the models' resolutions as "OBS_". The OBS_05

is in 0.5° resolution for CoLM, ISBA, JULES and LPJ-GUESS. The OBS_CLM4.5 and OBS_UVic are in the resolutions of CLM4.5 and UVic separately.

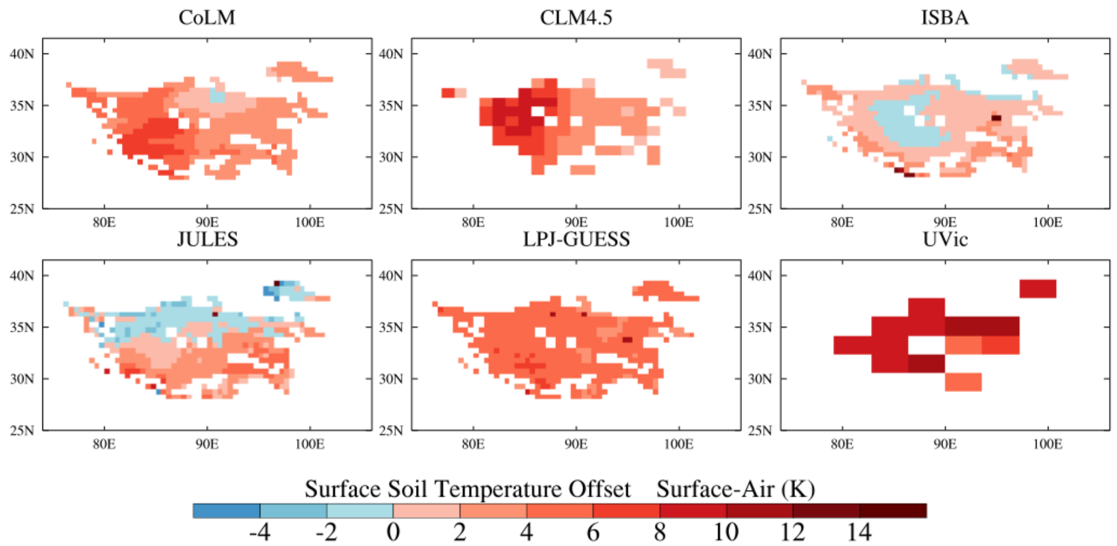


Figure 7. Mean surface temperature offset: difference in mean winter temperatures between surface soil and air, averaged over 1980-2000. Warm colors indicate soil is warmer than air temperature.

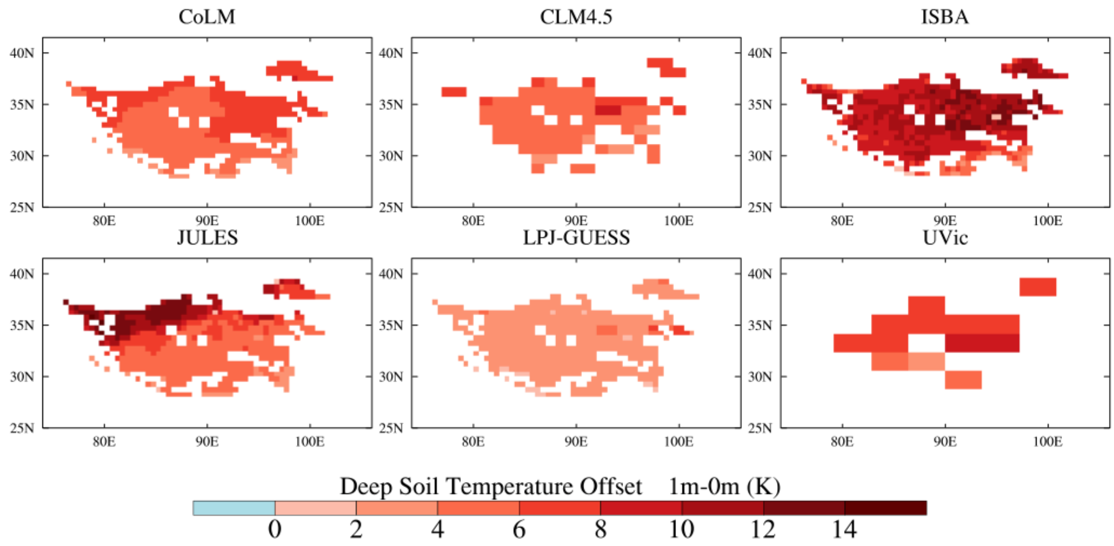


Figure 8. Mean soil temperature offset: difference in mean winter temperatures between soil at 1 m depth and surface soil, averaged over 1980-2000. Warm colors indicate deep soil is warmer than shallow soil.

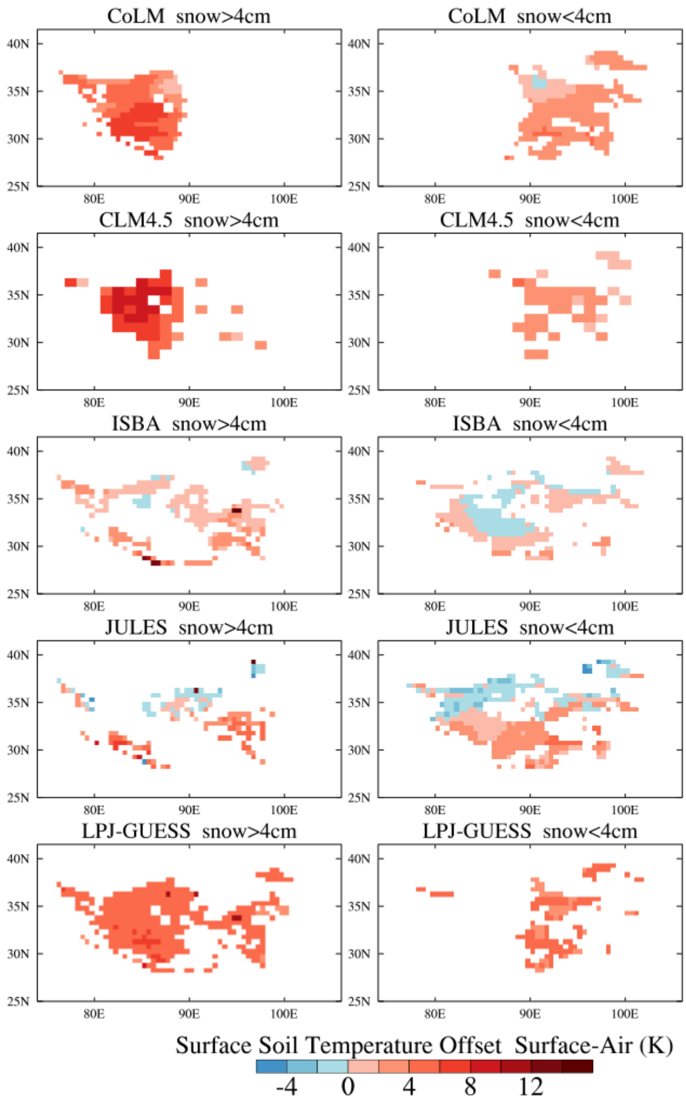


Figure 9. Mean surface temperature offset (difference in mean winter temperatures between surface soil and air, averaged over 1980-2000). Left column is for snow depth > 4 cm, right column shows regions with snow depth < 4 cm. Warm colors indicate soil is warmer than air temperature.

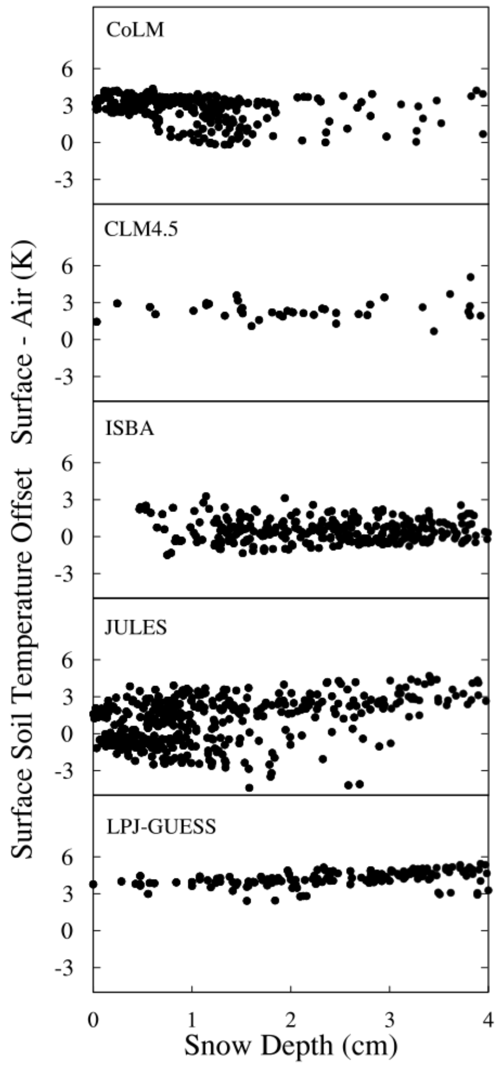


Figure 10. Mean surface temperature offset (difference in mean winter temperatures between surface soil and air, averaged over 1980-2000) as a function of snow depth for grid points where average snow depth < 4 cm.