Review to Wang et al. "Diagnostic and model dependent uncertainty of simulated Tibetan permafrost area"

This manuscript shows the uncertainties of model permafrost area estimates on the Tibetian Plateau (TP). It focuses on 5 different methods of permafrost calculation from 6 different model results. The MS is clearly written and the message is well communicated. The authors have addressed some important aspects of the cryospheric sciences, namely the TP and different calculation methods. The introduction is relevant and states the motivation and background of this study. The methods of permafrost calculation are well described. However, the discussion and conclusions need some more work. Further improvements can make this manuscript ready to published in TC. Please see my comments below.

Major comments / suggestions:

- 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. 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 snow depth data (days snowdepth > a threshold value like 1-5 cm). Then compare the air vs surface temperature offsets during this snow season and see the results.

- There needs to be a 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. Also some reference papers for each model should be included.

- 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 and why you choose to compare these sites. Especially at Fig.4, the 0.04 m observation of D105 and 2.63m observation of D110 are missing. You can better explain the reasons in a subsection. 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.

- 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.

- 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.

- 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. 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.

- To improve the scientific value of your model intercomparison results you have to tackle each of the following issues: 1.forcing data, 2.model spatial resolution, 3.model timestep, 4.model spinup, 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 (timestep) 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 Table1, you did not mention them in the discussions.
So you should clear the issues regarding to points 2,4, and 5

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

- 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 timeseries plots in Fig. 4 and Fig.5.

- Why is LPJ-GUESS always the coldest? That must be one simple process that is uniquely different than other models. 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 heat transfer, boundary conditions at deep soil, or the treatment of snow and vegetation cover and several other soil parameters. One obvious problem for this model's results is that why is it colder even though it has higher snow depth.

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

- 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.

- 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.

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

Minor comments:

- It would be better and more direct to avoid parentheses inside the abstract

P1771 L17 and L20: produce "better" permafrost maps of the TP you mean?
P1772 L9: lose the comma
P1772 L9: "plays" -> "play important roles"
P1772 L11: lose the comma
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…
P1774 L6: model's -> models'
P1775 L9: remain -> remains
P1775 L10: model studies -> model-based studies
P1775 L14: most of these models can provide daily temps or even sub-daily temps. You should a least mention the restriction of the model results that are available from RCN.
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.
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.
P1781 L3: lose the comma
P1781 L4: if you are talking about Cohen's paper, then you should put the reference out of parenthesis
P1781 L18: sites -> sites'
P1783 L24: I don't understand what you're 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.
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 Fig4a and Fig4c, almost all models (except LPJGUESS) 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 should revise this sentence. Yes in Table 4, we can see ISBA is the only

one that satisfies the <0.2C condition for all sites/depths but considering there is only 2 site for the lower depth (2.63), it is hard to generalize

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

P1785 L18: classed -> classified

P1785 L18: permafrost -> 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.

P1785 L21: revise the first sentence of sect 5. Too long to deliver the message clearly.

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

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

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

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

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.
- It doesn't make sense to put Wang06 map in between methods. You should make a separation between methods and the observational map.
- Can you also put description to the smaller two maps under the panel (Tibet and common region).
- Site locations are not very visible. Try to choose another marker and make them bolder

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

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

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.

Table 5: Start all words with capital letter. What does "snow cover: none" mean for UVic? No snow representation? This has to be mentioned because it affects everything for soil thermal dynamics… What do you mean by "unfrozen water effect during phase change"? Does that mean no freezing/thawing occurs in CoLM, LPJ-GUESS, and UVic?