We would like to thank Anonymous Referee #2 (AR2) for their constructive and positive comments. Below, we will respond to the comments made by AR2: the comments from AR2 in black, our response in blue.

The soil freezing and thawing processes are vital for land surface models because they influence both the soil thermal and hydrological variables and modified the land surface memory which play important roles in determining land-atmosphere interactions. However, current land surface models solve the freezing-thawing cycle by using a separated method, where soil temperature is firstly calculated and then soil freezing and thawing processes are adjusted. Li et al., coupled a physically more realistic and computationally more stable and efficient frozen soil module (FSM) into a land surface model the third-generation Simplified Simple Biosphere model (SSiB3-FSM), and evaluated the new model over the Tibetan Plateau and northern China systematically. They also investigated the influences of soil freezing-thawing cycle on the soil temperature profile, maximum frozen soil depth and soil memory. Generally, the work is important for the land model community and deepens our understanding of the influences of soil freezing-thawing processes. However, some minor revisions are still needed before its publication.

We thank the reviewer for a thorough and knowledgeable reading of the paper and your helpful comments and suggestions for the improvements of the manuscript. We have responded to the following comments or questions and modified the manuscript accordingly.

1. The author used the GHCN-CAMS product to evaluate the T2m simulations of SSiB3-FSM. I am confused that whether the temperature forcing from the Princeton is already the T2m? If so, how do you process the temperature forcings as land surface models usually need the forcings at the lowest level of atmosphere model?

R: It is true that surface air temperature is included in Princeton forcing data and is used to drive the SSiB3-FSM and SSiB3 in this study with other forcing variables (e.g. radiation, precipitation, wind and humidity). Princeton forcing data set provides a long-term, globally consistent data set of near-surface meteorological variables that can be used to drive models of the terrestrial hydrologic and ecological processes for the study of seasonal and interannual variability. This data set is constructed by combining a suite of global observation-based datasets with the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis. It has been widely used by many model studies. In this study, we aimed to validate the land surface skin temperature (not T2m) simulated by SSiB3-FSM and SSiB3 globally. So GHCN-CAMS surface air temperature data, which is a station observation based global monthly land surface air temperature data and was developed at the Climate Prediction Center, National Centers for Environmental Prediction, was used to validate the models in view of its superiority over other data sets over global domains. Although GHCN-CAMS data is air temperature data, in fact the changes in surface air temperature are highly consistent with those in skin temperature. Therefore, in this study Princeton forcing data was used to drive the models and

GHCN-CAMS air temperature data was used to validate the simulated land surface skin temperature globally. We added related information at line 207-210.

2. The soil layer depth of the SSiB3-FSM is 3m and the soil temperature stations over TP are all seasonally frozen ground whose maximum frozen depth is shallower than 3m. However, for some regions such as the western TP, the maximum frozen depth will be deeper than 3m. Will this influence the results in section 4.2.1 which seems to use the whole TP as study region.

R: The deepest soil column depth in SSiB3-FSM is not only 3m. The soil column in model is discretized into eight, eleven and twelve layers for desert, grassland and trees, respectively. And the deepest soil depth is also depends on the upper vegetation type. For example, the deepest soil depth over bare soil and grassland is 7.77 m and over forest it is about 12m. They are much larger than 3m. In this study, over TP we just focus on the 14 sites (Figure 1). They are almost located in the eastern part of TP, so the MFD is also shallower than western TP. We added the information about soil depth at line 128-130.

3. In section 4.2.2, the author said that land memory in TP is not given because it has been well studied by previous works. While I still suggest to give the results in the supporting information to improve the good performance of SSiB3-FSM.

R: Thanks for the suggestion. We added the land memory in TP in Table 3 in section 4.2.2.

4. The author compared the observed and simulated normalized maximum frozen depth in section 4.4, and I wonder why do you directly compare the original values?

R: We added the reason why we compared the normalized MFD in section 4.4. See line 458-460. Because there is a large difference between the original values for MFD in simulated results and the observation data, the zero-score normalization method was used to process the observed and simulated MFD. The difference of MFD between simulation and observation may come from the systematic bias between forcing data and observation. This bias is unavoidable and it does not affect the variation or trend of soil temperature, just shown in Figure 4 and Figure 5. Therefore, the zero-score normalization method is used in order to analysis and compare the variation and the trend of MFD both by simulation and observation.

5. In section 4.1.1 and Table 1, although the author give the evaluation results of T2m at global scale, the detailed results over TP and NC are still needed.

R: We gave the detailed results over TP and NC in section 4.1.1 and added them in Table 1.

6. The subsections in 4.1 should be revised. Change the “(2) Soil temperature profile in the TP” to “4.1.2 ...”, the “(3) Soil ...” to “4.1.3 ...”.

R: Yes. We have revised them.