Dear Reviewer,

We would like to thank you for your constructive and helpful comments, which helped us to improve our manuscript. Significant changes have been made according to your comments and suggestions.

The following is our point-by-point response. The reviewer's comments are shown in ***blue italics***. Our responses are provided in **black**. The revised text is in **red**.

Sincerely,
All of the authors

Reviewer #2

This paper examines how a set of state-of-the-art subseasonal-to-seasonal (S2S) fore- cast systems predict the Tibetan Plateau-wide snow cover and surface temperature over the 1999-2010 period. The forecast systems from ECMWF, CMA and NCEP are used in the intercomparison. The evaluation of forecasted snow cover is made against the multi-instrument (IMS) snow cover data. A connection is drawn between the bias is snow cover, which increases systematically with lead time in all models, and a bias in surface temperature. Model experiments with the WRF model support the idea that the latter is induced by the snow excess through land-atmosphere coupling. Few papers have examined snow forecast on the S2S time scale on the Tibetan Plateau (TP), a region with well-known biases in snow and surface temperature. This is an innovative study that is well-worthy of publication in The Cryosphere. I however recommend a major revision of the paper, before it is in an acceptable form for publication.

MAJOR COMMENTS

1) Some brief description of how the three different models are initialised in terms of snow and land surface is needed, complementing the description of the land-surface models. What sort of snow analysis is used to initialise the different models? Isn't ERA-5, which has strong snow biases over the TP, used to initialise the ECMWF S2S reforecasts? Which observations are used in these snow analyses, both globally and over the TP more specifically? While IMS is not used to *initialise the ECMWF forecast model over the TP (see Orsolini Y. et al., 2019), is it used in the other systems? The quality of this initialisation would most certainly influence the snow forecasts at the S2S time scale. The initial snow values may be a key addition to bring in Fig 3.*

Response:

Thank you for the useful comments. We completely agree with the reviewer that the information of snow initializations in S2S reforecasts has to be described clearly because they are the key factors influencing the TPSC prediction – the topic of our study. With a careful check, we confirmed that the northern hemisphere IMS snow cover data and ground observations of the snow depth available on the Global Telecommunications System (GTS) are used to initialize snow in the ECMWF S2S model. This is different from snow initialization in the ERA-5. Descriptions of snow initializations in the S2S forecasts were added in revised manuscript.

"For the ECMWF model, realistic snow is initialized in the forecasts. The snow mass is initialized by the ECMWF Land Data Assimilation System (LDAS). The snow initialization relies on a land surface synoptic report and national ground observations of the snow depth available on the Global Telecommunications System (GTS), as well as on the Interactive Multisensor Snow and Ice Mapping System (IMS) snow cover information. The NCEP model also initialize realistic snow in the forecasts. The snow initialization comes from the Climate Forecast System Reanalysis snow analysis using IMS and the Air Force Weather Agency snow depth analysis. Snow in the CMA model was not directly initialized in the forecasts. The initial conditions of the snow in the CMA model are from a balanced state produced by long-term air-sea initialization integration. See the details on snow initialization in the S2S models at https://confluence.ecmwf.int/display/S2S/Models." (in the revised Section 2)

2) Some more details about how the snow cover fraction conversion is derived in each model is needed. For example, a 100% snow cover may mean very different snow depth or snow water equivalent (the actual prognostic variables) in the different prediction systems. Also, I believe that IMS provides a binary information snow cover being 0 or 1, with the former case meaning that the fractional snow cover is below 50%. When aggregated to a 1-degree grid (L88), isn't there a range or uncertainty in the IMS aggregated value (given that the observed fractional snow cover could actually be 0 or 50%)? Please clarify these points and the implications for forecast verification.

Response:

1. We added detailed descriptions about how the snow cover fraction conversion is derived in each model in the revised Section 2.1.

"According to the snow scheme in each land surface model, we obtain the snow cover fraction, which is a diagnostic variable in this study.

The snow cover fraction (f_{snow}) in the ECMWF model is parameterized as follows:

$$f_{snow} = min[\ 1,\ S/(0.1 \times \rho)\ ] \tag{1}$$

where *min* indicates the minimum function, S is the snow water equivalent (unit is kg m^{-2}), and ρ is the snow density (unit is kg m^{-3}) (Dutra et al., 2010).

The f_{snow} in the NCEP model is parameterized as follows:

$$f_{snow} = min[\ 1,\ 1 - (e^{-0.001 \times 2.6 \times S/SNUP} - 0.001 \times S/SNUP \times e^{-2.6})\ ] \tag{2}$$

where e is the natural logarithm, and $SNUP$ is the vegetation parameter, which indicates the threshold snow depth (in water equivalent m) that implies 100% snow cover (Koren et al., 1999; Ek et al., 2003). The $SNUP$ ranges from 0.01 to 0.08 for different vegetation types. Details on the Noah code and vegetation parameters can be accessed in https://ral.ucar.edu/solutions/products/unified-noah-lsm.

The f_{snow} in the CMA model is parameterized as follows:

$$f_{snow} = min[\ 1,\ 1.77 \times d/(d + 10.6)\ ] \tag{3}$$

where d is the snow depth (unit is cm), which is calculated from the snow water equivalent and snow density (Wu and Wu, 2004)."

2. We used a method similar to Orsolini (2019) to aggregate a 24-km resolution IMS analysis, which is interpolated into the 1°×1° grid of the S2S models. Details are provided in revised Section 2.2.

"The original 24-km resolution IMS analysis is interpolated into the 1°×1° grid of the S2S models. IMS provides binary snow cover information: it has the value of 1 if more than 50% of the 24-km pixel is covered by snow; otherwise, it is 0 (snow free). Orsolini et al. (2019) aggregated the original IMS product to a lower resolution rectilinear grid. They counted the number of pixels with a value of 1 in a grid box; assuming that they have 100% cover gave the high estimate, and assuming that they represent 50% cover gave the low estimate. These two estimates provide a range of values, which reflects the uncertainty inherent to aggregating the 24-km binary data, e.g., a value of 1 in a pixel means a 50% to 100% snow coverage. Here, we used a method similar to Orsolini et al. (2019) to interpolate the original IMS product into the 1°×1° grid of S2S products, but we further averaged these two estimates."

3) The authors consider the snow cover averaged over the entire Tibetan-Plateau, but they do not show any snow cover maps although there is considerable heterogeneity, as shown by the authors in previous publications. I wonder if there could be compensation effects between different geographical sub-regions, that could result in an agreement of the TPSC index between forecasts and IMS. Different prediction systems may have different regional biases. Showing such maps would re-assure the reader that the prediction systems capture the main climatological features of the snow distribution over the TP and its S2S variability (e.g., in subregions as in Li W. et al., IJOC, 2019). Could spatial pattern correlation between IMS and snow forecasts be helpful in this case?

Response:

Good suggestions. We plotted additional figures in the revised manuscript to show the spatial pattern of the systemic bias (Fig. 4 in the revised manuscript), as suggested by the reviewer.

"Taking the differences between the forecasts with a lead of 4 weeks and the forecasts with a lead of 1 week as an example, the spatial patterns of these increases in the biases in the three models show some similarities (Fig. 4). Although the spatial patterns of the differences in the three models show some small discrepancies, the differences are mainly positive in the three models, especially over parts of central and eastern Tibetan Plateau. These indicate that the increasing TPSC with the forecasting lead time are at a regional spatial scale." (in the revised Section 3)

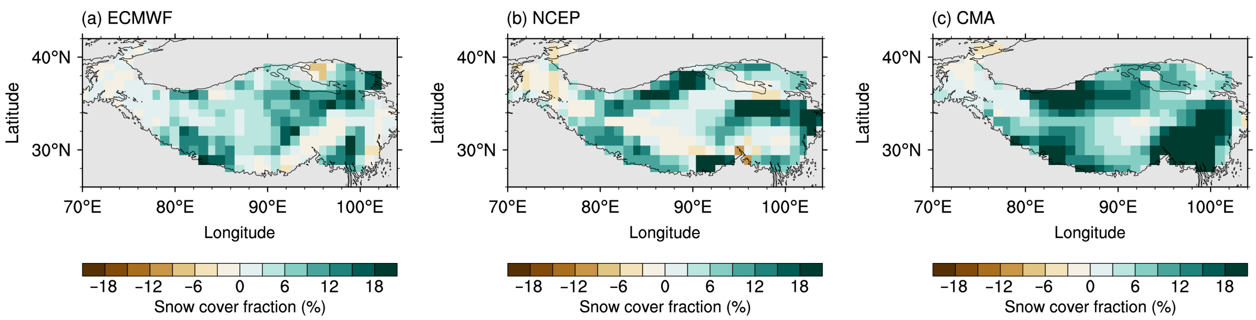

Figure 4 in the revised manuscript. Differences in the multiyear wintertime mean Tibetan Plateau snow cover fraction (unit: %) between forecasts with a lead of 4 weeks and forecasts with a lead of 1 week in (a) ECMWF, (b) NCEP and (c) CMA.

4) The authors could also try to examine the snow-temperature coupling strength (as an C2 indicator of land-atmosphere interaction) in the respective prediction systems by correlating the local forecasted temperature and forecasted snow, as a function of lead time. See Diro and Lin (2020) or Li et al. (2019).

Diro, G. T., and H. Lin, 2020: Subseasonal Forecast Skill of Snow Water Equivalent and Its Link with Temperature in Selected SubX Models. Wea. Forecasting, 35, 273– 284. Orsolini, Y., et al. G. (2019), Evaluation of snow depth and snow cover over the Tibetan Plateau in global reanalyses using in situ and satellite remote sensing ob- servations. The Cryosphere,Vol. 13.(8) s.2221-2239

Li, F., et al. (2019). Impact of snowinitializationinsubseasonalâĂŘˇtoâĂŘˇseasonalwinterforecastswiththeNorwe- gian Climate Prediction Model. Journal of Geophysical Research: Atmospheres, 124. https://doi.org/10.1029/2019JD030903

Response:

Many thanks for your good suggestion. The temporal correlations between snow cover fraction and SAT with forecast lead times of 1 week and 4 week for each grid point in three models were computed to identify the extent and nature of the snow-temperature relationship (Fig. 7 in the revised manuscript).

"The local SAT over the Tibetan Plateau is highly correlated with simultaneous TPSC at a subseasonal time scale (Li et al., 2020a). Local snow-temperature relationships in S2S models were examined. We took a similar approach as in F. Li et al. (2019) and Diro and Lin (2020). The temporal correlation between the snow cover fraction and SAT with a lead of 1 week and 4 weeks for each grid point in the three models was computed to identify the extent and nature of the relationship (Fig. 7). Almost all of the regions exhibit a significant negative correlation in all of these three models. Additionally, such a relationship in all three models did not weaken with the forecasting lead time (compare Fig. 7a–c and Fig. 7d–f), even if the forecasting skill on the TPSC declined over time. The reason is that the relationship between the snow cover fraction and the SAT is embedded in the land surface model." (in the revised Section 4)

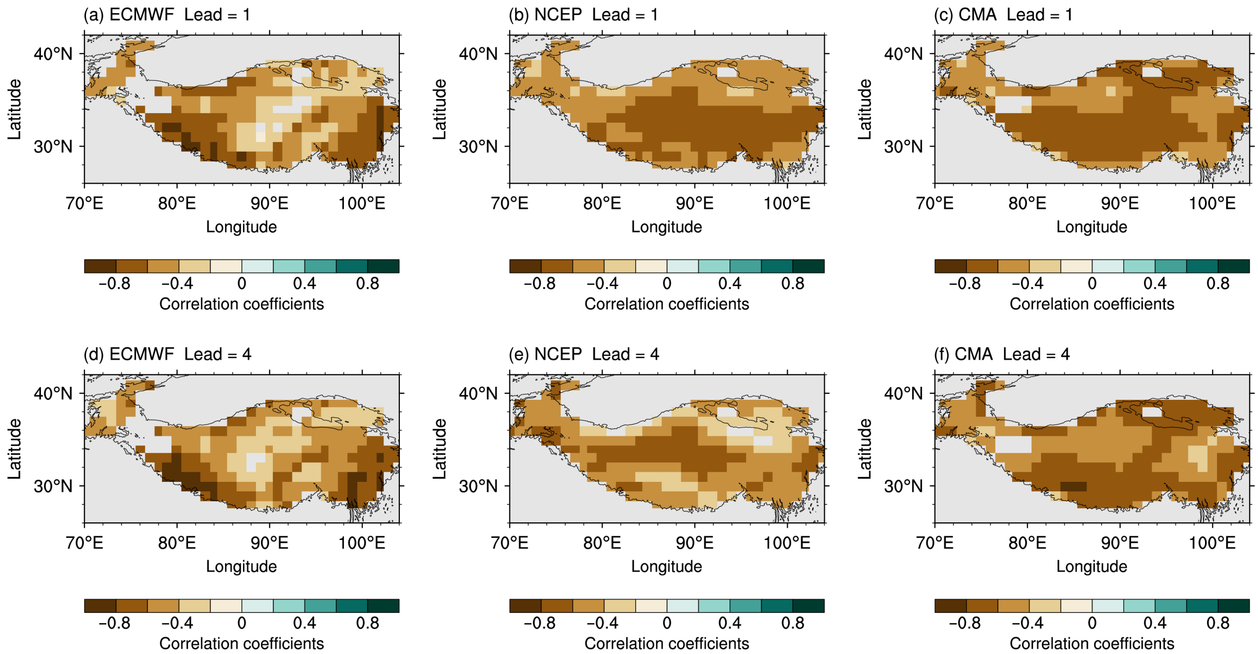

Figure 7 in the revised manuscript. Spatial pattern of correlations between the snow cover fraction and the surface air temperature with a lead of 1 week in (a) ECMWF, (b) NCEP and (c) CMA. Only significant correlations at the 0.01 level are displayed. (d)–(f) is similar to (a)–(c) but for forecasting with a lead of 4 weeks.

MINOR COMMENTS

L92-93: the description here focuses on winter. While winter is the main focus, many plots show year-long results. It would be better to emphasize the whole set used (nb of years, total nb of forecasts,. . .), not only the winters.

Response:

We are very sorry for the confusion. This issue is also raised by Reviewer #1. In fact, we attempted to focus on the TPSC study only for boreal winter considering that the TPSC biases during summer are not consistent among different models and there might have some complex processes involved. We have clarified the season that our present study focuses on and will leave the investigation of summertime TPSC prediction for our future work.

L252: the snow bias might not only arise from the land surface model (shared between WRF and the NCEP model) but also from the meteorological forcing (e.g., excess precipitation). Please clarify.

Response:

Motivated by this useful suggestion, we analyzed the precipitation in the forecast data and found that models tend to predict excess precipitation. The overestimated precipitation can explain why predicted snow cover increases with forecast lead times (Fig. 6 in the revised manuscript).

"Studies have shown that current state-of-the-art atmospheric general circulation models (GCMs) tend to strongly overestimate the precipitation over the Tibetan Plateau (e.g., Su et al., 2013; Chen and Frauenfeld, 2014; Zhang et al., 2016; Zhang et al., 2019). For example, Su et al. (2013) evaluated 24 GCMs that were available in the fifth phase of the Coupled Model Intercomparison Project (CMIP5) over the eastern Tibetan Plateau by comparing the model outputs with ground observations, and they found that all of the models consistently overestimated the observed precipitation for all seasons. Zhang et al., (2019) found similar results, in that all climate models they evaluated exaggerated the daily precipitation in the Tibetan Plateau during winter compared with the observed values. Here, we also found that the S2S models tended to overestimate the precipitation over the Tibetan Plateau. We compared the precipitation in the S2S models with both the gauges-based GPCC precipitation dataset and the satellite-based TRMM precipitation dataset (Fig. 6). The regional averaging wintertime mean precipitation over the Tibetan Plateau in the GPCC and TRMM models are 0.27 mm day^{-1} and 0.32 mm day^{-1}, respectively. Compared with the overserved precipitation, all three S2S models exaggerate the regional precipitation obviously. Notably, such exaggerations always exist throughout the model integration. The ECMWF model reproduces the precipitation that is closest to the observations among the three models, but it still shows a large overestimation. The precipitation in the ECMWF model is 0.78 mm day^{-1} to 0.88 mm day^{-1}. The precipitation values in the NCEP model (1.07 mm day^{-1} to 1.37 mm day^{-1}) and in the CMA model (1.50 mm day^{-1} to 2.13 mm day^{-1}) have larger precipitation biases and even increase with the forecasting lead time. These overestimations of the precipitation induce underestimations of the TPSC dissipation, and they lead to positive biases in the TPSC from the models. Because the overestimation of the precipitation exists throughout the model integration, the positive biases of the TPSC accumulate and increase with the model integration." (in the revised Section 3)

"All of the three S2S models consistently exaggerate the precipitation over the Tibetan Plateau compared to the observations. The exaggeration of the precipitation is prominent and always exists throughout the model integration. Systematic bias in the TPSC therefore occurs and accumulates with the model integration time due to exaggeration of the precipitation in the models." (in the revised Section 5)

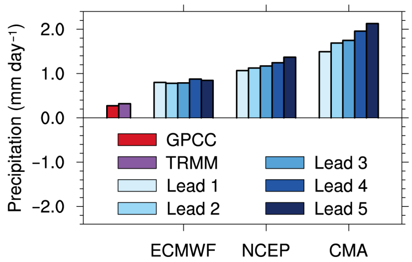

Figure 6 in the revised manuscript. The multiyear wintertime mean precipitation over the Tibetan Plateau (unit: mm day^{-1}) for the observations and forecasts in each model.

L244: how is the GDAS snow analysis used in Section 4.3 on numerical modelling compare with the IMS snow analysis, used in the first part of the paper. While it is mentioned that GDAS assimilates IMS, does it assimilate it over the TP? What does it assimilate specifically over the TP (in-situ data?)? Would the prediction skill be different if evaluated against GDAS (Fig 1) ?

Response:

The main conclusion in this study is that there is a systematic positive bias in the TPSC, and this bias increases with forecast lead times. To support this conclusion, the multiyear winter mean TPSC derived from the S2S models are compared to the observations.

Following your suggestion, we checked the multiyear winter mean TPSC in the GDAS (the FNL analyses). The multiyear winter mean TPSC index is 28.5% in the FNL analyses, and 31.9% in the IMS analyses. Compared to the systematic bias of TPSC in S2S models presented in this study, the difference of winter mean TPSC index between the FNL and the IMS are much smaller and do not change the main conclusion.

Conclusions: a brief mention of possible, relevant physical processes over the TP leading to snow ablation would be helpful. Could it be the strong surface winds or else the snow sublimation missing in the models? The short length of the period over which the forecasts are evaluated (around 10 years) is a bit of a concern. It appears that the biases are quite strong and systematic, but I wonder if some features in the forecasts would be robust over a longer period: for example, the slowdown in TPSC in early winter (December) seen in ECMWF and NCEP (Fig 3). I realise that if adding another 10 years may entail a lot of computational work, but it would add to the robustness of the conclusions. At least, a word of caution in the summary is warranted.

Response:

1. In the conclusion, we ascribed the sources of the systematic bias of TPSC prediction could come from the biased precipitation in the models. We agree with the reviewer that the

surface winds and snow sublimation could also play a role in causing the snow ablation. Identifying the relative contributions of these factors to the biased snow prediction needs more detailed and careful diagnoses. A discussion was added in Section 5. "Surface winds and snow sublimation could also play a role in causing the snow ablation. Identifying the relative contributions of these factors to the biased snow prediction needs more detailed and careful diagnoses. Future studies on this issue are potentially valuable." We will address this issue in the near future. Thank you for your comments.

2. We conducted another 10 years of numerical experiments. The experiments in the revised manuscript ran for 20 winters (from 2000/2001 to 2019/2020).

Typos / English

L43: hydrologic cycle

L28: radiative rather than radiant

L108: the total variability

L113: and the three different

L143: the preceding week rather than the last week, seems more appropriate

L152: accumulation, leading to a systematic TPSC bias.

L168: growth is used for a declining variable. Either decrement, reduction or decline should be clearer.

L173: real rate should be observed rate or rate derived from IMS.

L173: indices or index

L202: growth of SAT, rather than TPSC (it says 1.2 degree).

L229: land-atmosphere interactions

L256: Hence,

Response:

All of these are corrected. We thank the reviewer for your careful reading of the manuscript.