Geoscientific Model Development Discussions Open Access EGU

Review of gmd-2020-329 by Xue et al.
"Impact of Initialized Land Surface Temperature and Snowpack on Subseasonal to Seasonal Prediction Project, Phase I (LS4P-I): Organization and Experimental design"

This paper introduces the LS4P project including its motivation, goals, and instructions on the modelling experiments. The key idea is to study the effect of soil temperatures in mountainous areas on subsequent precipitation in downstream areas through remote effects of land-atmosphere interactions. Models and observations hint that such effects could be existing, and could consequently be exploited for weather/climate forecasting. Anaccurate representation of these effects across models is challenging, and the project is aiming to improve this.

Recommendation:
I think the paper requires major revisions.

This paper comprehensively presents the LS4P project which brings together the land- climate modelling community. The key idea about exploiting remote effects of land- atmosphere interactions for weather predictability is promising, and fits well with other recent studies illustrating so-far largely overlooked effects of land surface status and fluxes in downstream areas. Also considering soil temperature in this context is innovative.as it also reflects to some extent the moisture/energy state of the land surface, and quite some satellite and ground-based data are available which are partly insufficiently exploited, particularly in comparison with the more prominent soil moisture. Nevertheless,I also see some shortcomings in this paper which should to be addressed to make the paper suitable for publication in Geoscientific Model Development:

Response: The reviewer has very carefully reviewed the manuscript, and provided very insightful, constructive, critical, and encouraging comments and detailed suggestions. The manuscript has been revised based on the comments/suggestions. We sincerely appreciate the reviewer's efforts and have acknowledged the reviewer and editor in the revised manuscripts.

(1)
I have some doubts about the application of the mask as described in section 3.2.While equation 1a is clear, I do not understand why equation 1b is needed;
in this context also lines 403-408 and Figure 2 are unclear to me.
More generally, I think that the models' memory is a dynamic feature which should be addressed through adapting the modelled (soil temperature and moisture) *dynamics*rather than their *states*.
This way, I feel if the tuning parameter "n" could exaggerate initial correctionswhich could degrade the simulations in the early forecasting period
Why not simply correcting the model/forecast biases through post-processing and without interferring with the actual model simulations?
Finally, as the mask correction requires observed soil temperature information, maybe I missed that but I was wondering how this is done in places where this is notavailable?

Response: The reviewer raises several issues for clarifications of our soil temperature initialization methodology. The following is our responses.

1). The LS4P project pursues a new approach, i.e., using the LST/SUBT anomaly in high mountain areas, to improve the S2S prediction. The current start-of-the-art models, however, are unable to properly produce the observed anomalies, and then by extension, this anomaly-induced dynamic and the associated physical processes, in their simulations. As such processes are not existed in the model simulation; the bias correction in post-processing is unable to generate these processes. In fact, the LS4P deals with the S2S prediction, which is essentially the same as weather forecasting. A bias correction in the post-process is normally not employed.

2). It is a good idea to improve the model dynamic and physical processes to overcome the modeling problem. However, improving Earth system model/land model and reanalyses data (which problem is shown in Figure 4 of this paper) in order to overcome the deficiency in modeling high mountain land surface temperature anomaly is not a simple task and may take decades of effort (today's land temperature model development has more than 70-year history), but proper S2S prediction, including drought/flood prediction, is an urgent World Meteorological Organization (WMO) task with societal implications. On line 377 of the revised manuscript, we pointed out that preliminary research suggests that "prescribing both LST and SUBT initial anomalies based on the observed T-2m anomaly and model bias is the only way for the current ESMs to reproduce the observed May T-2m anomalies". This is current status in the LS4P modeling groups. Of course, we welcome any research group to find a new way to produce observed anomaly, but it is not a simple task. On other hand, using initialization to improve meteorological prediction is nothing new but a traditional meteorological approach. We believe when this issue gets more attention and more data from measurement are available, more methodologies may be developed. But the scientific development always takes time. We have to undertake any development by making step-by-step improvements.

3). The reviewer has raised question about the schematic diagram (Figure2), which is designed to help other modeling groups to reproduce the mask that we use to test our approach. We acknowledge that the presentations of the Figure 2, its caption, and relevant text were not comprehensive and had shortcomings in helping readers to understand our approach. We have had several iterations among our co-authors and revised the figure, the caption, and the related presentation in the text to help readers better understand our approach as well as the procedure to produce the mask for initialization. In particular, we have clarified in the revised Figure 2 which initial temperature is for Task 1 and which one is for Task 3 with more details and explanations in figure note. We also reorganize Section 3.2(3). We believe these should help readers to understand the overall ideas in this figure and in the LS4P-I experimental design.

Meanwhile, the original Figure 2 includes both warm and cold years. In the revised version, to make the thing simple and less confusion for readers who are unfamiliar with LS4P, we only include cold year (same as the case in the LS4P Phase I) in the text and move the schematic diagram for the warm case to appendix for readers who want to pursue this issue further for their own research.

4). Regarding the observational data, under the current big data system, it should be available very soon. The general public can obtain last month's observational data. In a major climate center, the real time analysis data are generated in a very timely manner. In terms of the model bias information, if it is unavailable in some cases, then the model climatology bias can be applied. Our paper (Section 5 and Figure 7) indicates that the bias is very consistent if comparing climatology and year 2003. The methodology does not require a precise bias value. But the general signs (positive or negative) are important.

5). The reviewer thought we may impose artificial large forcing because of a tuning parameter "n". It is not true. In the following paper in a Climate Dynamics special issue, we will show every model's-imposed forcing. They are not that extreme. The LS4P-I includes most of major climate centers in the world. If our approach is totally different from their normal practices, they would not endorse the LS4P-I and participate in this project. By the way, many parameterizations, such as convective schemes, aerosol parameterizations, etc. also have the tuning parameters.

(2)
The description and motivation for selection of the ground truth data is insufficientin my opinion. The text in lines 249-273 lists many datasets but does not indicate how they are applied. A summary table of all employed datasets would be nice, includingtheir spatial and temporal extent, advantages, and variables provided.
I understand that you are using the CMA dataset as this is based on a relatively largenumber of ground measurements. This makes sense, but it would be interesting how(well) this dataset extrapolates between these measurements and how many (fewer?)measurements are employed by other datasets such as the state-of-the-art ERA5 reanalysis. Next to this, I was wondering why you are not employing available satellite-basedland surface temperature products http://data.globtemperature.info/ ?

Response: The data sets described in this section are produced by the LS4P-I data group, and we intend to provide this information for scientists who are interesting in conducting further LS4P-I researches. These measurements are the foundation for the LS4P-I research. For instance, in the reviews of earlier papers/proposals of the LST/SUBT effect, some reviewers just use "there are no large scale SUBT measurements in high mountain areas to confirm the exist/presence of such anomaly" to suggest for rejection. Now the LS4P-I data groups have provided comprehensive data sets to support the community for the research. That's why we choose the Tibetan Plateau as the focus area for the LS4P Phase I and would like to provide relevant information for the community to introduce these data sets. In the revised manuscript, per the reviewer's comments, the role of the data group is clearly presented (lines 186-187; 251-254). We have also added a table in the appendix (Table B1) listing all the data that we used.

As to the station number in China, for other data sets, such as ECMWF data set, they normally obtain the data from GTS system and through collaborative agreements. Normally, they have less than 400 stations, much less than what we have listed in the text.

Remote sensing over the Tibetan Plateau is challenging because of a lack of validation data (for testing the algorithms). Since the Tibetan Plateau is one of the focus areas of GLASS satellite products, the GLASS group has more experience on this area's remote sensing with better quality there. In addition, the GLASS group is participating in the LS4P-I project. That's why we mainly use the GLASS products in the project. But we certainly do not exclude other satellite products and will use them if they can provide useful information for the project.

Thank you for sharing the link (http://data.globtemperature.info/) for surface temperature data. Unfortunately, all these datasets have a short coverage period and do not cover the climatology period (1981-2010 and 1980-2013), which we have considered in our study.

(3)
The manuscript is comprehensive but also quite long. To make it more concise it would behelpful to shorten where possible I think. Below, I have indicated some examples where content is repeated and where I would see potential to shorten the text.
Moreover, a summary table listing the main information regarding tasks 1-5 in section 3.2would improve the readability.
Response: Thanks for the reviewer's suggestion. We will make the revision according to every suggestion that you list in specific comments (see our response in Specific Comments below for detail). In the revised manuscript we have added a table (Table 1) to clearly provide information's regarding Tasks 1-4 as suggested.

(4)
The authors refer a lot to snow effects in sections 1 and 2, and I like these ideas.However, snow is not mentioned at all in sections 3, 4 and 5, and apparently only implicitly (through LST) part of

the analyses.
This should be clarified, and the role of snow in the project as described insections 1 and 2 should be toned down.

Response: The snow part of the work is an important component in LS4P. Snow is one of the major driver to produce LST/SUBT anomalies. That's why it is included in the project's name. In Phase 1, however, we are mainly looking for first order effects most related to the soil surface and deeper layers; but indeed, we mentioned snow effects as we (agree with this reviewer) think it is important and related to the LST/SUBT anomaly, and plan to examine it in the Phase II. Since this is the main paper introducing the LS4P, we have to present the importance of snow in this project in section 1. Otherwise, the readers may immediately raise issue about where the LST/SUBT anomalies come from.

The shortcoming in our previous presentation was not to provide a clear expectation how much snow related activity will be discussed in this paper, and make some readers keep waiting till the end of the paper. We now make a clarification at the end of Section II that this part of research will be considered in Phase II papers and will not be presented further. So readers (who are interest in this) will have a proper expectation.

(5)
As with the snow, also predictability is prominently mentioned in sections 1 and 2 and even in the abstract and title, but the detailed description of the project and thesimulations does not refer at all to this. So also here I would suggest to either include details on how the predictability could/will be assessed, or tone it down in the beginning of themanuscript.
Response: Please see our response above.
I do not wish to remain anonymous - René Orth.
Response: Thanks. We have acknowledged the reviewer in the revised manuscript.

Specific comments:
line 29 & 179: "Data groups" is not clear.
Response: Sorry, we failed to clearly provide the relevant information. In the response to your main comment 2, we had a discussion on this issue. In the revised paper, we have clarified their contribution to the LS4P on lines 186-187; 251-254.

line 35: Summer precipitation in which area?
Response: We have added "beyond East Asia" on line 35

line 49: "stubbornly low", not everywhere, there are quite some regional variations ofprecipitation forecast skill
Response: The S2S prediction is associated with drought/flood/heatwave prediction. In N. America, Europe, West Africa, and East and South Asia, the Earth System Models have difficult to reproduce these events. That's why the WCRP and the WRRP/WMO list S2S as the current high priority. In some areas, such as California, the weather and climate have low variations in some seasons. But this is not the focus of our S2S prediction research.

lines 60-80: in this context you could cite Orth and Seneviratne (2017) where we comparethe impacts of SST vs soil moisture for land climate globally
Response: Added and thanks for the suggestion

lines 83-85, and elsewhere: I am missing some justification why you chose to focus on (usually data sparse) mountain areas, and why their downstream impacts are expected tobe higher than that

of flat regions with possibly larger surface heat fluxes.
Response: Lines 85-111 (the revised version) in the introduction and the 1st paragraph in Section 2 discuss the justification to use high mountain areas as focus for S2S prediction. The LST's effect in high mountains have been long overlooked until scientists in the LS4P groups conducted preliminary researched discovering its important role in S2S predictability. The preliminary sensitivity experiments and relevant papers presented in this part show the effect of spring LST/SUBT in high mountain areas on downstream summer precipitation. The papers cited here also discussed the mechanisms why the impacts are in downstream. Because of the high elevation, the perturbations induced from the land surface process could propagate for long distances through Rossby wave interactions affecting the downstream areas. We have modified the paper in the introduction to help the reader to better relate the discussions there to the justification as to why we choose a high mountain area as the focus.

Yes, the mountain area measurements are overlooked because of the difficulty making measurements in the environmental condition there and because of failing to understand the importance of the measurement in these areas. We hope our research will stimulate more measurements in these areas.

line 86: "very close", you mean strongly related here?
Response: On line 89 of the revised version, the text has been modified to clarify the meaning here.

line 89 & Figure 1 caption: I do not fully understand how this was computed. Do youselect years with warm/cold Mays, respectively, and then you track the temperaturedifference between these years through all months?
Response: Yes, we use the approach as you mentioned. This approach has been pointed out in the revised note after the caption of figure 1. We further emphasize it on line 95 of the revision

line 100: SSiB, abbreviation not explained
Response: Added

line 122: insert "the" before "snow darkening effect"
Response: done

lines 127/128: could you please specify/explain "large diversity" a bit more?
Response: we have modified the sentence and in the revised version of the manuscript it reads as "this could be one of the major reasons for the large discrepancies in simulated T-2m and its anomaly in current Earth System Models (ESMs)".

Line 146: insert "the initiative" after "historical development of"
added

lines 178/179: there is no need to put both written-out and numeric numbers there
Response: agree. We have eliminated the written-out numbers.

lines 195/196: I think this is an important part of the project, can you give some moredetails on this data base?
Response. We have added relevant information for the data base on line 203 of the revised version.
lines 210/211: Sorry for mentioning an own study again, but Orth and Seneviratne (2017)can be instructive here I think
Here just summarize questions, no any citation listed here. The relevant citations are in Introduction and the reviewer's paper has been added in citation there.
lines 221-225: repetition, could be removed line 246: insert "the" before "Tibetan Plateau"
done

line 248: please be more specific which of these datasets are useful for the project, andwhy

The data used for this project is clarified on lines 249-251 with a table in Appendix B in the revised version. This section mainly presents data sets produced by the LS4P data groups. Only some of these data are used by the LS4P phase I experiment. Most data introduced here can be used for the LS4P related research, such as the causes of LST/SUBT anomalies, Tibetan Plateau land surface characteristics (for instance, snow, frozen soil) associated with land memory and land surface energy and water balances, and land/atmosphere interaction there. The multi-model testing in the LS4P project can only investigate some key issues and intends to stimulate more research on this aspect from the community, which is needed to ultimately understand this issue. This section provides the relevant data information for the community, which is very useful for them to conduct high mountain-related researches. We have clarified this in the revised section 3.1 of the manuscript.

lines 267-269: could be removed

As indicated earlier, during the early LST/SUBT research, this was the precise reason some reviewers reject our approach. We make the statement here is to show although some of the data do not directly use in LS4P phase I experiment, but they do provide basic information/evidence to support the crucial justification for the LS4P activity.

lines 282-283: you mean anomaly precipitation rates of +1.32 mm/day?

Response: Yes. Anomaly is added.

line 294: "around late April", why is this not more specific?

Response: The LS4P requests for at least 6 members for each Task, which normally means 6 different starting dates (for initial conditions), such as April 25, April 26, April 27, April 28, April 29, and April 30. Every modeling group may select different days based on their normal practice for numerical prediction (for instance, some days from April 26-May 1, and some from April 28-May 3). That's why we use "around late April" here.

line 316: "the models' performances are then checked" can be removed

done

line 321: what is a "proper" lapse rate?

Response: This issue is discussed in Xue et al. (1996) and Gao et al. (2017). We modify the sentence to make it clear.

lines 323-327: repetition, can be removed

Okay.

lines 334-335: Not sure if the project is obsolete if models would do a good job, as you could still study LST/SUBT downstream effects, as stated in the project goals in e.g. the abstract

Response. The downstream effect is a new scientific discovery for the S2S predictability regardless whether the models are able to produce the LST anomalies on the mountain areas. However, if the models were able to produce proper temperature anomaly over the Tibetan Plateau, the focus would not be how to generate the observed LST anomaly. We have modified the statement on line 344 and add lines 624-625 to make it more adequate and precise.

line 349: "around 2010", why is this not more specific?

Response: Every climate modeling center has made long term climatological runs but with slightly different starting and ending years. We are unable to request these big centers to redo their climatological simulation because of the huge amount of work load and computer time, so as a best approximation we only ask them to send their climatology which is an average around certain time period in order to be compatible with our experiment. We believe a climate based on the average 1980-2010 and another from 1979 to 2009 should have no fundamental difference.

line 426: what are the "sensitivity" and "control" runs?
Response: The sensitivity is Task 3 run and control is Task 1 run. We have clarified this.

line 450: here you could point to Table S1
Done. Thanks.

lines 478-479: ERA-Intermin should be updated to ERA5 (https://climate.copernicus.eu/climate-reanalysis)
We did not update due to the time period that ERA5 covers. The GMD also does not allow to put website address in the paper, so we still use the previous citation.

lines 473-475: I get the point that you would use the dataset that uses most ground station measurements as reference, but I am wondering how many stations ERA5 is using over this area?
Response: I do not know exactly how many stations ERA5 has. But based on my information (pers. Comm.) it should not be more than 400 stations in China, which is much less than what we use.

lines 590-591: While it becomes more clear later, it would be good to already motivate here why you are comparing biases with anomalies.

Response. This is a good suggestion. We have added lines 624-625 to indicate why we want to compare bias and anomalies.

line 594 & Figure 6: I do not understand why you do this multiplication with -1.
Response: Results in Figure 6 provide the underpinnings for the LS4P conjecture: if the May land temperature anomaly on the Tibetan Plateau does contribute to the June East Asian precipitation anomaly, then improving the May land surface temperature simulation over the Tibetan Plateau through an improved initialization should make Earth System models to produce better June East Asian precipitation.

For instance, May 2003 was a cold month for the Tibetan Plateau and June 2003 was dry to the south of the Yangtze River in the observations. If we postulate the May cold T2m in the Tibetan Plateau caused the drought, then a model with cold (warm) bias in May in the Tibetan Plateau should produce a dry (wet) bias to the south of the Yangtze river if these Earth System models' dynamics and physics reflect such linkage that in the real world.

Because some models have a cold and dry bias and some have a warm and wet bias, when we make the composite in Figure 6, we have to multiply "-1" for the models with warm/wet bias to integrate them with the models with cold and dry bias (to avoid their biases cancel each other), and to compare them with observed anomalies.

line 605: How is this correlation computed? Is it a spatial correlation? Over which domain?
Response: it is a spatial correlation over the figure domain, which now is specified in the revised version (lines 638-639).

line 615: Please specify the area over which the subsequent drought occurred.
It is clarified on lines 647-649

line 637: Why these methods, and not a similar approach as for example in Koster et al. (2016)
Response: Here we deal with the prediction. Smith et al.'s work on this line of the original manuscript is a statistical prediction. Koster et al.'s work (2016) conducted series of stationary wave model (SWM) experiments in which the boreal summer atmosphere is forced, over a number of locations in the continental United States, with an idealized diabatic heating anomaly. As such, Koster's work is an ideal sensitivity study and is different from what we pursue

(prediction study).

lines 643-647: Why and how is the fore-restore method causing inaccurate soil memory?
Response: We have two published studies (Liu et al., 2020; Li et al., 2021) addressing how the force restore method causes the problem and how to improve the memory through the proper parameterizations. Since a comprehensive discussion on these issues are out of the scope for this paper, in the revised text we added these two citations on line 680.

Liu Y., Y. Xue, Q. Li, D. Lettenmaier, and P. Zhao, 2020: Investigation of the variability of near-surface temperature anomaly and its causes over the Tibetan Plateau. J. Geophy. Res. 125, e2020JD032800. https://doi.org/10.1029/2020JD032800.
Li, Q., Xue, Y., and Liu, Y.: Impact of frozen soil processes on soil thermal characteristics at seasonal to decadal scales over the Tibetan Plateau and North China, Hydrol. Earth Syst. Sci., 25, 2089–2107, https://doi.org/10.5194/hess-25-2089-2021, 2021.

line 666: Correct tense, 2020 is in the past now :)
Corrected

line 669-670: "A possible ... will also be prepared" can be remove
Done

Figure 5:
- the quality/resolution is very low, please improve
Done

References:

Orth, R. and S.I. Seneviratne
Variability of Soil Moisture and Sea Surface Temperatures Similarly Important for Warm-Season Land Climate in the Community Earth System Model
Clim. Dyn. 30(6), 2141–2162 (2017).

Koster, R., et al.
Impacts of Local Soil Moisture Anomalies on the Atmospheric Circulation and on RemoteSurface Meteorological Fields during Boreal Summer: A Comprehensive Analysis over North America
J. Climate 29(20), 7345-7364 (2016).