Geoscientific Model Development Discussions

Geosci. Model Dev. Discuss., referee comment RC2
https://doi.org/10.5194/gmd-2020-329-RC2, 2021

Comment on gmd-2020-329

Anonymous Referee #2

Referee comment on "Impact of Initialized Land Surface Temperature and Snowpack on Subseasonal to Seasonal Prediction Project, Phase I (LS4P-I): organization and experimental design" by Yongkang Xue et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2020-329-RC2, 2021

This paper describes the international / multi-organization LS4P project and provides some initial analyses of data submissions. It is essentially an "introduce the community to the project" paper, with most of the scientific findings to be documented later, as the project progresses.

While the paper is well written, in my opinion it has enough issues to rate a review of "major revision". On the one hand, I do applaud the experiment organizers for bringing together such a wide-ranging group of participants to address such an important problem. With this diverse a group and a corresponding collection of model outputs from such a diverse set of models, I don't doubt that the project will bear useful scientific fruit, and a paper like this that introduces the project to the broader community is certainly of value. On the other hand, the paper's write-up glosses over several critical aspects of seasonal and subseasonal prediction that call some of the project's long-term strategies into question, at least in terms of how they're currently described. A revised version should address these shortcomings through substantial qualification (not just a sentence or two here and there) or, better yet, through a substantial rethinking of the approaches to be applied.

1. The underlying assumption of the project appears to be that if a model does not produce an accurate temperature over the Tibetan Plateau, the fault lies with the land model (e.g., in how long the land model maintains an initial condition). The paper states this explicitly on line 465. The truth is, all models have biases in both air temperature and precipitation across the globe, and these biases could have any number of sources. A temperature bias over the Tibetan Plateau might have nothing to do with land model processes. It might instead result from deficiencies in the reproduction of the general circulation, for example, or from some problems with the radiation balance. Forcing the model to have a low temperature bias by imposing a stronger initial temperature anomaly (perhaps even an unrealistic anomaly, through eq. 1) may amount to "getting the right answer for the wrong reason", which is not a good basis for a forecast experiment. It's quite possible that forcing a correct temperature through such an initialization when the model wants to do something else for reasons unrelated to land processes might have unexpected negative consequences – especially if the model is artificially modified in one region and not in surrounding regions. Substantial discussion regarding this is needed.

2. The overall strategy seems to ignore the fact that forecast models generate their forecasts relative to their own climatologies. A model that is known to be biased warm may produce an anomalously cold 2003 over the Tibetan Plateau (compared to what it

usually produces there), and that would be useful information even if the forecasted temperatures are still warmer than the average observed TP temperature. The point is that people know how to account for long-term model biases. They would properly consider a forecast model's result to be "2003 will be colder than usual by 5 degrees" rather than "the temperature will be 20C". The emphasis here on matching the observed temperature in absolute magnitude seems inappropriate to a discussion of forecast systems. See, e.g., the NMME forecast anomaly pages [https://www.cpc.ncep.noaa.gov/products/NMME/seasanom.shtml], which show forecasted anomalies relative to each model's climatology.

3. Forecast systems also produce a range of forecasted values through the running of ensembles, and any one ensemble member could represent what happens in nature. The experimental analysis protocol, however, emphasizes the importance of having the *ensemble mean* match the observed anomaly. This is inappropriate. The key question is, do any of the ensemble members look like the observations? (And, in conjunction with point #2 above, the truly key question is, do any of the ensemble member anomalies *relative to the forecast model's climatology* look like the observed anomaly?) A model cannot be considered wrong if one of its ensemble members looks like the observations. Insisting that an ensemble mean match a specific year's temperature seems wrong.

4. The model results concerning May Tibetan Plateau temperature anomalies and June precipitation anomalies in east Asia is perhaps suggestive but far from indicative of a causal relationship. Even if the agreements in 6a/6c and 6b/6d do suggest that one pattern led to the other (it could very well be coincidental), I don't see how the Tibetan Plateau in 6a/6c can be isolated as the source of the agreement in 6b/6d. Significant qualification of this figure's implications is needed.

Minor comments:

-- Just to clarify: Are the warm and cold years the same for each month shown? If not, it's not clear what this figure says about the persistence of warm and cold anomalies (line 92).

-- lines 106-108. This sounds strange. Can the authors clarify the link between temperature and water amount? While I see that more water leads to a slower seasonal transition, it's not obvious that at a single point in time, temperature tells you something about water amount.

-- I studied Figure 2 for a long time and still can't make sense of it. Why, for a cold year initialization, does the initial condition for a model with a cold bias get set to climatology whereas that with a warm bias does not? Also, please clarify in the caption: are the biases discussed here errors for the particular year of simulation, or are they long term climatological biases? I'm guessing the former, since Task 2 would need to be done for the latter; in that case, though, the use of the term "bias" is confusing here. Bias should refer to a long-term climatological error (reflecting a model deficiency), not to the error at a specific time (which should reflect both bias and random error). Overall, Figure 2 is not helpful for explaining the approach. And again, based on my earlier comments, I'm not convinced the Tmask strategy is appropriate anyway.

-- Equation 1 appears to be a means to impose an artificially large temperature anomaly at the start of a simulation so that the anomaly is maintained realistically during the forecast. As far as I can see, there's no physical basis for the equation; it's fully empirical and could lead to initial temperatures that make little physical sense (e.g., colder than the model ever gets). More qualification is needed regarding how artificial this construct is. (And again, based on my earlier comments, it may not be appropriate to fix the temperature error in this way, since it may have a source other than the land model.)

-- line 211 (and elsewhere): Replace "SST" with "ocean state", since SST is only a small part of what seasonal forecast systems rely on from the ocean. Arguably, subsurface ocean temperature distributions are more relevant.

-- Clarification regarding figure 7: is this the average of the 2003 anomalies relative to the different models' climatologies, or is it the average (over all years) model T-2m and precipitation minus the average (over all years) observations? I'm guessing the latter, given the remarkable agreement with figure 6cd. The latter can truly be considered a bias, but the term bias was used differently elsewhere in the paper.