The study investigates predictions of surface air temperature under strong and weak L-A coupling conditions using NASA's GEOS S2S forecast system. The authors assess the predictive skill of temperature forecasts during weeks 3 and 4 of the boreal summer, particularly in the Midwest and northern Great Plains, by applying L-A coupling metrics. These metrics capture the relationships between soil moisture, latent heat flux, and some other atmospheric variables, such as surface skin temperature and planetary boundary layer height. The study hypothesizes that surface temperature predictions improve when strong coupling is present. The results show that strong L-A coupling increases the ability to predict extreme warm events, especially for the northern Great Plains, where strong coupling increases the likelihood of correctly predicting abnormally warm temperatures during weeks 3-4.

While the paper is well-written and concise, the discussions are limited and could be expanded to address several key issues. For example, it should be addressed how some preprocessing explained in the method section, such as upscaling or spatiotemporal aggregation of the different datasets to match each other, plays a role in the results. Moreover, the strength of land-atmosphere coupling is mostly mediated by soil moisture, especially in water-limited regions. Despite its importance, there are no discussions about the accuracy of the soil moisture used in this study. Even a small bias in soil moisture values used in this study, especially in heavily irrigated regions during the growing season, may have a significant impact on the subseasonal air temperature predictions. Providing a comparison with observational-based soil moisture observations such as SMAP would more clearly identify the regions where the strength of land-atmosphere coupling is more reliable in contributing to air temperature prediction skills.

Thank you for your comments. Below we address your two key points: 1) the impact of the spatial interpolation on our results and 2) the accuracy of the soil moisture.

1) The impact of spatial interpolation

Motivated by your comment, we conducted the same analysis using GEOS-S2S-2 data and ERA5 on a 0.5°x0.5° grid (without spatial interpolation for GEOS-S2S-2). Result, presented in Fig. R1, are similar to what we found in the original analysis. Accordingly, we have added the following sentence at line 218:

"To test whether our spatial interpolation of the GEOS-S2S-2 data had some impact on our findings, we conducted the same analysis (not shown) using a 0.5°x0.5° grid without the spatial interpolation. The results were essentially the same."

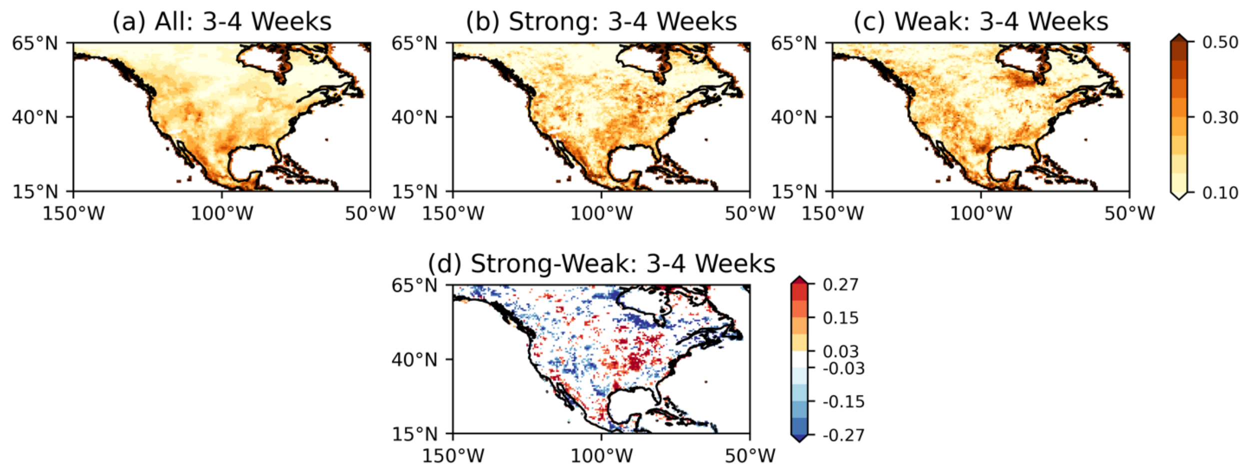

Figure R1. The ACC of T2m anomalies (forecast values versus observations) at a forecast lead time of week 3-4 over North America during (a) all, (b) strong, and (c) weak L-A coupling events. Strong (weak) events are defined when all three indices are above (below) the 50 percentiles. (d) The difference in T2m ACC between strong and weak events. Statistically significant differences at a 90% confidence level are shaded.

2) The accuracy of the soil moisture

We agree with the reviewer; even a small bias in soil moisture values used in this study, especially in heavily irrigated regions during the growing season, may have a significant impact on the subseasonal air temperature predictions. As background for the manuscript's new text about this, we should point out that there is no fully satisfactory way of evaluating the soil moisture accuracy. SMAP data exist from only from 2015 onward, allowing for only a small overlap period; furthermore, like all satellite-based products, the SMAP data have their own biases. Assimilation-based soil moisture datasets (GLDAS, NLDAS, etc.) are model-dependent and thus subject to model assumptions, and the in-situ measurement sites that provide soil moisture are point measurements that, due to the spatial representativeness problem, may not accurately describe the grid-scale soil moistures that our study relies on.

All this being said, we have added the following text to the paper at line 76 in response to the reviewer's comment:

"The weakly coupled data assimilation system used for the forecast system's initialization makes use of observed (rain gauge-based) precipitation measurements to drive the land surface fields; as a result, the soil moisture initialization for the forecasts appropriately reflects the character of the antecedent observed precipitation. Quantitative estimates of the accuracy of such precipitation-driven, model-based estimates are rare but do exist. Using a slightly modified version of the land model used here in GEOS (Catchment model), Reichle et al. (2017) found strong temporal correlations between the surface and root zone soil moisture so generated and corresponding in-situ measurements across the contiguous US (CONUS). The soil moistures produced by a slightly updated version of the land model were recently compared to Soil Moisture Active Passive (SMAP) satellite-based soil moisture retrievals (Entekhabi et al., 2009); the resulting anomaly correlation coefficients over CONUS were found to generally exceed 0.6 (Qing Liu, personal comm., 2024).

Koster et al. (2020, their Figure 9) effectively show that significant subseasonal air temperature forecast skill in the GEOS-S2S-2 system, particularly in the eastern half of CONUS, is strongly tied to the initialized profile soil moisture, indicating useful accuracy therein."