Reviewer comments
Authors responses

Reviewer 2

This study probabilistically quantifies the patterns of the sink-to-source carbon transition in the northern high latitudes under various emission scenarios (including the overshoot scenario) by means of the PRIME-JULES framework that explicitly couples permafrost physics, dynamic vegetation and fire processes. This study could fill the gap that key carbon cycle processes have not been jointly coupled in the current CMIP6 models.
The reviewer correctly summarises the paper and identifies the gap we are trying to fill, we thank the reviewer.

As an ESM emulator, the PRIME framework can only simulate effects of climates on terrestrial carbon cycle, and fails to simulate the feedback of carbon cycle to climate system. That is, it fails to capture the two-way carbon-climate coupling effects. Furthermore, it does not couple the biophysical feedbacks of terrestrial ecosystems to climate (e.g., the impacts of vegetation albedo and evapotranspiration changes on regional temperatures), which may lead to biases in the simulation of long-term carbon cycle feedbacks.
We agree this is a limitation of using land surface models offline in this way, opposed to the computationally expensive coupling into an Earth system model. However, using the PRIME framework has allowed us to investigate the carbon cycle dynamics including processes yet to be fully coupled in ESMs, i.e., permafrost and dynamic vegetation; processes which we have shown to be especially important in the northern high latitudes. Additionally, PRIME allowed us to produce a large ensemble size compared to what would be possible with an ESM, enabling the sink-to-source transition to be investigated on a much more granular level. This limitation of our framework is highlighted and addressed in the Conclusions section, where we state:
"*Additionally, simulations with coupled models will allow for an evaluation of the ecosystem carbon balance including the additional biophysical feedbacks, a current limitation with the PRIME framework.*"

The northward expansion of dynamic vegetation in the model is only driven by climate and CO_2, with no incorporation of factors such as soil nutrient limitations (excluding nitrogen), physical constraints of permafrost instability on plant growth, and human activities (e.g., forest management). In the real world, these factors can significantly inhibit the northward expansion of boreal forests, which may result in an overestimation of the enhancement effect of dynamic vegetation on the carbon sink.
The reviewer correctly identifies processes that are not currently incorporated in land surface models, which will also influence the northward expansion of vegetation. This is discussed in the manuscript in the following text:
"*In reality there are additional sources of uncertainty which determine the net carbon balance, such as the amount of boreal forest expansion that will be seen (Rotbarth et al. 2023). For example, there are uncertainties in the amount of tree growth potential within permafrost ecosystems, where it is argued that the allocation of nutrients within a tree will prioritise remaining upright on unstable grounds rather than growth in these regions (Alfaro et al. 2024). The spatial patterns will also be influenced by complex interactions with water and nutrients which could limit growth potential (See et al. 2024; Zhao et al. 2022), and additionally, the impact of fire creates large amounts of uncertainty to the spatial carbon loss (Virkkala et al. 2025).*"

Simulations in this study are solely based on the single land surface model JULES. The model results have not been validated with in-situ observational data (e.g., long-term carbon flux observation stations in tundra and boreal forest regions), leaving the generalizability of the results to be verified.

The results from this study are based on a single land surface model and the importance of additional land surface models including high northern latitude processes to enable inter-model comparison is highlighted in the conclusions:

"*Investigating the balance of carbon fluxes in this region in additional land surface models will allow for inter-model comparison and help to further understand the sensitivity of such fluxes to climate change (Padron et al. 2022).*"

However, it is noted that the novel methodology presented here is the large ensemble which provides a probabilistic analysis of the net land carbon cycle flux, which has not been previously possible with a model such as JULES. This work demonstrates the potential to expand to other land surface models used in ESMs.

Additionally, the manuscript already included a section in the methods which evaluates the JULES-pf configuration (see section 2.3). In this section, JULES-pf is evaluated against CARDAMOM v14 (an observationally-informed model-data fusion analysis), and we additionally compared contemporary NEP estimates with both top-down and bottom-up approaches presented in Hugelius et al. 2024 (see lines 147 to 154). We also highlight the large uncertainty reflected in the data with this value.

"*Hugelius et al. 2024 used three broad methodological approaches to quantify the NEP: land surface models, atmospheric inversion and in-situ upscaling (including JULES-ES and CARDAMOM V13), each of which has a large uncertainty range. The total plausible range of NEP spans from a net source of -0.66 to a net sink of 1.37 Gt yr^{-1}; the large range reflects data uncertainties, including whether fire is explicit or implicit, and methodological uncertainties.*"

The 95% percentile range of CARDAMOM appears to be overly broad, and thus it is understandable that the results of JULES-pf fall within this range. However, this cannot prove that the results of JULES-pf are reliable or superior to those of other models. Please clarify what the 95% percentile range of CARDAMOM indicates.

The evaluation of JULES-pf is not to imply superiority over other models, it is to use CARDAMOM's observationally informed uncertainty to qualify JULES-pf as reliable. The evaluation assesses output variables against observationally-informed values to evaluate its performance with respect to the information carried in current state-of-the-art geospatial datasets. We do not evaluate against field-based estimates due to the mismatch in spatial scales (observation <1 km^2, JULES simulation ≅200-300 km^2). Whereas satellite-based Earth Observation and upscaled geospatial datasets give the opportunity to evaluate at an appropriate spatial resolution. However, these datasets have been shown to be inadequate to evaluate land surface models as different products that estimate the same ecological variable (e.g. GPP, AGB) do not agree with one another often several times their reported uncertainties (e.g., Seiler et al., 2022).

Our solution is to evaluate using an observationally-informed systemic analysis, enforcing consistency between the observational datasets. On this basis the evaluation shows that JULES-pf's contemporary performance is plausible. For detailed description see Section 1.1 in Supplementary Information. In brief, CARDAMOM is a Bayesian calibration framework applied at pixel-scale, thus a local analysis, to retrieve ensembles of parameters consistent with observations and their uncertainties for an intermediate complexity terrestrial ecosystem model (DALEC) providing a data-informed diagnostic analysis. The 95 % confidence interval is directly estimated from the pixel-level ensemble of parameters and the resulting DALEC simulations.

Thus representing a realistic uncertainty range as informed by a variety of datasets (inc. biomass, leaf area, absorbed photosynthetic radiation, soil C, geospatial priors on plant traits).

Seiler, C., Melton, J. R., Arora, V. K., Sitch, S., Friedlingstein, P., and Anthoni, P..: Are terrestrial biosphere models fit for simulating the global land carbon sink? Journal of Advances in Modeling Earth Systems, 14, e2021MS002 946, 2022

Line 64: Mathison et al. (2025) should be cited in the correct reference format.
Corrected.