Reply to Referee comments on egusphere-2025-6393: „Unravelling the tree cover dynamics over the last 20,000 years on the Northern Hemisphere” by Anne Dallmeyer et al.

Reviewer 1 : Qiong Zhang

Overall assessment

This manuscript presents a comprehensive and ambitious evaluation of Northern Hemisphere tree-cover dynamics over the last ~20 kyr, combining a transient MPI-ESM1.2 simulation with the recently developed hemispheric REVEALS-based tree-cover reconstruction by Schild et al. (2025). The study goes well beyond a descriptive model–data comparison by systematically diagnosing spatial patterns of agreement and disagreement, disentangling climatic drivers using emulation and GAM approaches, and explicitly discussing structural limitations of current dynamic vegetation models (DVMs).

The manuscript is clearly written, methodologically sophisticated, and addresses questions that are highly relevant to the paleoclimate community, particularly those working at the interface of palaeodata synthesis and Earth system modelling. The use of a transient simulation rather than time-slice experiments is a major strength, as is the explicit treatment of non-linearity in climate–vegetation relationships.

Overall, I find this to be a strong and publishable contribution, but in its current form it would benefit from clarifications, tighter framing of some conclusions, and a more critical separation between (1) climate biases, (2) vegetation model structure, and (3) reconstruction limitations. My comments below are intended to strengthen the robustness and interpretability of the results rather than to challenge the major findings.

A: We thank Qiong Zhang for her thorough, well-structured, and constructive review of our manuscript. Her thoughtful comments and suggestions have helped us to significantly improve the clarity and quality of the paper. Our detailed responses to the reviewer's comments are provided in the following. For clarity, the reviewer's comments are shown in black, and our responses are given in blue.

1. Interpretation of REVEALS tree cover versus modelled absolute cover

A central issue in this manuscript is the comparison between REVEALS-derived tree cover (which sums to 100% vegetation) and MPI-ESM absolute tree cover including bare ground. The authors correctly acknowledge this mismatch and justify their choice to compare absolute PFT area (fi) rather than relative fractions (ci).

However, this choice has far-reaching implications for the interpretation of MAE, variance differences, and the systematic bias pattern (overestimation at low tree cover, underestimation at high tree cover). At present, these implications are discussed mainly qualitatively.

I suggest that the authors

(1) add a concise conceptual clarification (possibly a schematic or boxed explanation) explicitly explaining how REVEALS tree cover should be interpreted in open landscapes, and how this affects MAE and variance metrics;

A: REVEALS-based forest cover represents the percentage of forest taxa relative to the total vegetation within the pollen source area. As a proxy, pollen primarily provides information about vegetation composition. The only way to infer the total amount of vegetation is through changes in pollen concentration in sediments over time; however, this metric is rarely recorded.

Uncertainty in REVEALS estimates increases when very large unvegetated areas are present, such as ice sheets, gravel fields or beds, water bodies, or deserts. Water areas were corrected for in the dataset under the assumption that they remained roughly constant through time. If other unvegetated areas are present, the REVEALS estimate of forest cover would exhibit a consistent positive bias. This potential direction of error is considered in our discussion of model–proxy mismatches.

We added to the REVEALS method part:

LL132: “In general, REVEALS-based forest cover is characterized by a slight overestimation. This arises partly from persistent taxon-specific biases and partly from the presence of unvegetated areas. REVEALS-based vegetation cover reflects the composition of vegetated areas but does not capture the overall extent of vegetation. As a result, forest cover can be overestimated when large unvegetated areas occur within the pollen source area.”

(2) Clarify more explicitly that part of the diagnosed non-linear bias pattern (Fig. 6) is methodological rather than purely ecological or model-structural, especially in sparsely vegetated regions;

A: We thank the reviewer for this helpful comment. We clarified in the manuscript that part of the non-linear bias pattern shown in Fig. 6 arises from the methodological setup, as the difference metric (MPI − REVEALS) is analysed as a function of REVEALS tree cover. This mathematical dependence can contribute to the apparent trend across vegetation classes. We therefore emphasise in the revised text that the figure should primarily be interpreted as illustrating the magnitude of model–data deviations along the vegetation gradient rather than a purely ecological or model-structural signal:

We wrote (LL318): “Part of this pattern, however, is methodological, because the difference metric (MPI − REVEALS) is analysed as a function of REVEALS tree cover. This mathematical dependence tends to produce increasingly negative values at high REVEALS fractions even in the absence of ecological effects. The figure should therefore primarily be interpreted as illustrating the magnitude of model–data deviations along the vegetation gradient rather than a purely ecological non-linear response.”

(3) Consider whether at least a sensitivity comparison using relative cover (ci) for selected well-vegetated regions (e.g. mid-Holocene Europe) could help bound the uncertainty. This clarification is important because many readers may otherwise interpret MAE patterns too directly as "model error".

A: We agree that a sensitivity comparison using relative cover (ci) could in principle provide additional insight into the uncertainty. However, in practice, such an analysis is complicated by several factors. Notably, land use is not represented in the model setup, wetlands are not included, and the model relies on fixed parameters (e.g. for disturbance strength), which can lead to differences between MPI-ESM simulations and REVEALS reconstructions (Dallmeyer et al., 2023), even when relative vegetation fractions are compared. Therefore, mismatches between REVEALS and MPI-ESM cannot be unambiguously attributed to differences in the treatment of vegetation cover (e.g. the fact that REVEALS reconstructs 100% vegetation cover, whereas MPI-ESM includes a bare soil fraction).

To provide at least a first-order estimate of the potential overestimation of tree cover in REVEALS, we instead included a comparison with MODIS-derived tree cover in the original manuscript. This offers an independent benchmark and helps to contextualize the magnitude of the mismatch.

We also note that there are emerging approaches to estimate the bare soil fraction within the REVEALS framework, for example using the modern analogue technique (Sun et al., 2022). However, these methods have so far only been applied to limited regions (e.g. part of China) and are not yet readily transferable to a global analysis.

We added to section 2.3 (LL144):

„These discrepancies between MODIS and REVEALS tree cover may provide a first-order indication of the magnitude by which REVEALS could overestimate tree cover due to the reconstruction of vegetation composition only. However, this potential bias is unlikely to be temporally constant, as it depends on changing environmental conditions, vegetation dynamics, and the extent of non-vegetated surfaces.“

2. Mid-Holocene forest maximum: model limitation or forcing limitation?

A key result is the failure of MPI-ESM to reproduce the reconstructed mid-Holocene tree-cover maximum across large parts of the Northern Hemisphere, with the model instead peaking in the early Holocene. This is an important and robust finding. However, the manuscript currently blends several possible explanations such as overly strong warm-season temperature control linked to insolation, missing processes (permafrost, soils, disturbances), climate biases (e.g. absence of a mid-Holocene thermal maximum), and possible reconstruction artefacts.

A: The temporal mismatch in tree cover maximum is indeed very prominent and is not just a feature of the boreal latitudes. It has to be further analysed. Therefore, we plan to focus on this topic in a follow-up paper since additional model experiments have to be performed to evaluate this mismatch and to assess the consequences of different tree cover dynamics, which would overload the paper in its current version.

We added this in the summary of the revised version (LL682):

„Consequently, the model is not able to capture the mid-Holocene forest maximum observed across large parts of the Northern Hemisphere. Instead, tree cover peaks during the Early Holocene in the model due to the strong sensitivity of the simulated tree cover to warm season temperature, which closely follows the insolation forcing. However, the underlying mechanisms and consequences require further analysis and will be addressed in a follow-up study."

I suggest sharpening this discussion by more explicitly separating (a) deficiencies in simulated climate trajectories (e.g. lack of MH warmth or hydroclimate persistence) from (b) deficiencies in vegetation sensitivity to that climate.

Also clarify whether the emulator experiments indicate that even with a corrected climate, JSBACH would still fail to produce a mid-Holocene maximum in boreal regions (which would strongly support a structural vegetation limitation). And explicitly position this result in relation to other transient Holocene simulations (e.g. PMIP-style experiments), even if only qualitatively.

A: Thanks for raising this point. The emulations using bias-corrected climate indeed fail to reproduce the mid-Holocene maximum in boreal latitudes. We attribute this primarily to the simple bias correction applied in this study. The delta method corrects for the difference between the modelled and modern climate at a single time slice and applies this constant anomaly across the entire 20,000-year period. This approach assumes a stationary offset, i.e., it cannot change the temporal evolution of climate and thus the period when temperatures peak. Changes in the temporal evolution of emulated vegetation for bias-corrected climate data is therefore purely a result of non-linear climate-vegetation relationships. Therefore, the timing shifts slightly in small areas but this approach does not change the hemispheric-scale temporal progression. Thus, this approach can only exclude that persistent biases in the modelled climate are responsible for the timing mismatch. Consequently, it is not possible to generally separate the effects of deficiencies in the simulated climate trajectory from those arising from limitations in vegetation sensitivity, unless dedicated sensitivity experiments are performed or independent climate reconstructions are used to adjust the forcing.

To further elaborate on this point, we have made the following revisions/clarifications in the manuscript:

LL547: "The model appears to miss the prolonged early- to mid-Holocene warming, pointing to deficiencies in the simulated climate trajectory."

LL552: "Neither simulation nor emulation reproduce the reconstructed Holocene increase, pointing to a wrong driver or structural vegetation deficiencies in the model."

LL 580: "Therefore, the mismatch in the timing of the Holocene tree cover maximum persists across large parts of the Northern Hemisphere in the emulation with bias-corrected climate, particularly in boreal latitudes. This can partly be expected because the applied bias correction removes only stationary offsets.

At the same time, the results indicate that the absence of a mid-Holocene vegetation maximum is unlikely to be explained solely by biases in the simulated climate. Instead, they point toward structural limitations in the representation of boreal vegetation dynamics and associated land–surface feedbacks in the model, which the emulator does not correct for."

LL667: "Our study also shows that the climate–vegetation relationship can regionally be well captured but that systematic climate biases can still produce large mismatches between simulated and reconstructed tree cover. Two mechanisms may contribute to these discrepancies. First, deficiencies in simulated climate trajectories can lead to incorrect climate forcing of the vegetation model. In the temperate forest–steppe transition zones of Central North America and East Asia, the model is systematically too wet, enabling tree growth much earlier than observed and causing major offsets. Bias correction can reduce the mismatch to the reconstructions here. At high northern latitudes, the model is too cold and warm season temperatures peak during the early Holocene, which distorts the simulated climate forcing of tree cover and contributes to unrealistic vegetation dynamics, e.g. in Western Canada.

Second, limitations in vegetation sensitivity to climate may further amplify these mismatches. In combination with the global and fixed bioclimatic thresholds used in the vegetation model, the cold bias likely alters the tree cover sensitivities to temperature in JSBACH (and thus in the emulator trained on it). As a result, bias correction does not improve the agreement with the reconstructions in all boreal regions, but instead increases the model-data mismatch, e.g. in Siberia, indicating a misrepresented climate forcing and an unrealistic representation of the boreal forest dynamic. Missing processes in the model, such as the impact of permafrost on the tree cover dynamic, further increase the model-data differences. Consequently, the model is not able to capture the mid-Holocene forest maximum observed across large parts of the Northern Hemisphere. Instead, tree cover peaks during the Early Holocene in the model due to the strong sensitivity of the simulated tree cover to warm season temperature, which closely follows the insolation forcing. However, the underlying mechanisms and consequences require further analysis and will be addressed in a follow-up study."

We have furthermore expanded the discussion to explicitly position our results within the context of transient Holocene simulations. We now explicitly discuss this finding in relation to previous Holocene modelling studies, which have highlighted both the importance of vegetation feedbacks for Holocene climate evolution and the persistent model–data discrepancies in high-latitude temperature trends. Several studies have suggested that insufficient vegetation–albedo or land-surface feedbacks may contribute to the difficulty of models in reproducing proxy-based mid-Holocene warmth in northern high latitudes (e.g.,Hao et al., 2025; Liu et al., 2014; Thompson et al., 2022). We have added a paragraph in the discussion section summarizing these results and placing our findings in this broader modelling context:

LL587: „This interpretation is consistent with findings from transient Holocene simulations. Several studies highlight the importance of vegetation feedbacks for Holocene climate evolution, in particular mid-Holocene warmth in the Northern Hemisphere (Thompson et al., 2022). However, models with dynamic vegetation frequently underestimate the magnitude and spatial extent of vegetation changes inferred from proxy records ((Dallmeyer et al., 2019; Harrison et al., 2015; Tian and Jiang, 2013), potentially limiting the strength of these feedbacks.

Transient Holocene simulations also show strong regional differences in the timing and magnitude of the Holocene Thermal Maximum, particularly in boreal regions. While models agree reasonably well in regions strongly influenced by residual ice sheets, substantial model–proxy discrepancies persist in areas such as Alaska and Siberia (Zhang et al., 2017, 2018). In these regions, simulated temperature trends during the Early Holocene often differ from pollen-based reconstructions (i.e. warming in reconstructions, cooling in models), highlighting persistent uncertainties in both climatic forcing of vegetation and regional feedbacks (Hao et al., 2025)."

3. Interpretation of CO_2 dominance and linearity

The manuscript convincingly shows that strong linear alignment of tree cover with CO_2 is associated with high temporal correlation but inflated variance and MAE, leading to the conclusion that CO_2 sensitivity is likely too strong in this model version. This is an important point, but it requires careful wording to avoid over-interpretation. The strong correlation between REVEALS tree cover and CO_2 is plausibly a proxy effect reflecting hemispheric deglacial trends rather than a physiological signal. This distinction should be emphasised more clearly in the Results and Discussion. It would be helpful to explicitly state that the emulator diagnoses model-internal sensitivities, not real-world sensitivities, and that the mismatch with REVEALS may arise from both sides. The conclusion that "CO_2 sensitivity is too strong" should be framed as relative to reconstructed variability patterns, not as an absolute statement about palaeo-CO_2 fertilisation.

A: We fully agree that this is misleading. We phrased it more nuanced in our revised manuscript and differentiate the pseudo CO2 relationship in the proxies from the probably too strong model-internal sensitivity:

LL364: “We emphasize that the emulator diagnoses sensitivities intrinsic to the model itself, and thus characterizes model behaviour rather than real-world sensitivities.“

LL405: „This statistical agreement may therefore primarily reflect the shared large-scale deglacial trend in both variables, rather than a direct plant-physiological CO_2 response in the real-world.

Because the emulator diagnoses model-internal sensitivities within the MPI-ESM framework, discrepancies between simulated and reconstructed patterns do not necessarily indicate model failure; they may also arise from characteristics of the reconstructions, including their temporal resolution, chronological uncertainties, and spatial aggregation, which can enhance alignment with long-term trends while smoothing local variability.“

4. Regional interpretation of disagreement: local dynamics versus model failure

The analysis of high- vs low-agreement grid cells is one of the strongest parts of the manuscript. The conclusion that poor agreement often coincides with non-linear, summer-temperature-dominated responses is compelling. However, I encourage the authors to make clearer that poor model–data agreement does not automatically imply "model failure", but may indicate regions where local ecological processes, migration lags, disturbance regimes, or microclimates dominate. This distinction is particularly important for Siberia, forest–tundra ecotones, and forest–steppe transition zones.

A: This is indeed an important point. Of course, the reconstructions are far from perfect. We decided against explicitly discussing the shortcomings in the reconstructions, as this has already been done in detail in studies on REVEALS reconstructions. However, we share the reviewer's opinion that it is important to point out possible sources of error in the reconstructions and incorporate this more specifically in the revised version.

We restructured the paragraph addressing the reasons for model data differences in section 3.2:

LL295: „These regional discrepancies between the model and the reconstructions point to several key challenges in comparing the REVEALS-based reconstructions with vegetation models. Possible reasons include methodological caveats such as shortcomings in the process of reconstructions, including the REVEALS model (Githumbi et al., 2022; Kaplan et al., 2017; Li et al., 2023; Schild et al., 2025), biases in the simulated climate, missing or idealized model components (Dallmeyer et al., 2022, 2023; Li et al., 2025; Zhang et al., 2025) and differences in the vegetation-climate relationship. These factors may lead to local ecological processes, migration lags, and disturbance regimes affecting the reconstructions, while being inadequately represented in the model. This may be particularly problematic in ecotones, where small climatic changes or disturbances can cause large shifts in vegetation composition because multiple vegetation types coexist near their ecological limits.“

And we added the following note to section 3.4 (LL428):

„It should be noted that high MAE values and low correlations do not necessarily indicate deficiencies in the climate-vegetation relationship in the model. They may also reflect the fact that the reconstructions capture local ecological processes, migration lags, or microclimatic conditions that are not represented in coarse-resolution simulation. Since these factors cannot be filtered out, all grid cells are included in the analysis to examine the dependence on climatic conditions.“

And we added to the summary, part c) (LL636):

„We note that low model–data agreement does not necessarily indicate model deficiencies, but may also reflect local processes mirrored in the reconstructions. Moreover, our analysis indicates that grid cells with high model-data agreement...“

Minor comments and technical suggestions

P5, Line 149, "We consider only grid cells that indicate a significant ($p<0.1$) correlation between the simulated and reconstructed tree cover." The choice of $p < 0.1$ for correlation significance should be briefly justified, especially given the large number of grid cells.

A: We used a relaxed significance threshold ($p < 0.1$) to retain grid cells across the full range of environmental conditions in the study area. This approach ensured that regions with sparse data availability and particularly low tree cover were not disproportionately excluded, as such areas tend to show weaker statistical significance due to lower signal-to-noise ratios despite potentially meaningful correlations between reconstructions and model data. Using $p < 0.1$ therefore helped to maintain spatial representativeness in the analysis and avoid a bias toward regions with higher tree cover or denser data coverage.

We changed the sentence in the manuscript to: “We consider all grid-cells where the correlation between reconstructions and model data was significant at $p < 0.1$ in order to maintain spatial representativeness, particularly in regions with sparse data availability and low tree cover where correlations tend to be less statistically significant due to lower signal-to-noise ratios.” (LL160)

P4, Line 116, it states that both model output and reconstructions are binned into 500-year intervals, but this temporal aggregation is not discussed further. Given that many key diagnostics (variance differences, non-linearity from GAMs, emulator performance, and correlation strength) are sensitive to temporal smoothing, the authors should briefly discuss how the 500-year binning may influence the detected strength of non-linear responses and the interpretation of model–data agreement, particularly in regions with rapid postglacial changes.

A: We thank the reviewer for pointing this out. The 500-year binning was applied to harmonise the temporal resolution of the REVEALS reconstructions and the model output and to reduce noise arising from age uncertainties and uneven sampling in the pollen records. While such aggregation may smooth short-term variability and reduce the expression of rapid postglacial changes, this effect is not expected to substantially influence our main findings, which are based on multi-millennial trends. Because the same binning was applied consistently to both model output and reconstructions, the comparison between them remains internally consistent.

We added to the method part (LL117):
“Such temporal aggregation inevitably smooths out short-term variability and may dampen rapid vegetation changes, particularly during periods of abrupt postglacial transitions. As a consequence, the magnitude of submillennial variability, correlations during abrupt centennial-scale climate changes, and the detectability of sub-millennial non-linear responses may be somewhat reduced. Therefore, the comparison focusses on differences in multi-millennial trends and sensitivities, instead of submillennial vegetation changes.”

P12, 372-373, and P18, Line 553, the use of "energy-limited" vs "water-limited", would be helpful to briefly define these terms (e.g. dominant driver in emulator/GAM sense) to avoid confusion.

A: This was indeed confusing. We already define them a few sentences before, but mixed up „water-limited“ and „water-driven“. We changed this in the revised version:

LL411: “The dominant climatic drivers in MPI-ESM according to the emulation vary through time. While energy limitations (i.e. Tw or CO2 as most influential variables) are a key driver during the deglaciation, moisture availability (i.e. precipitation and soil moisture dominant driver) becomes increasingly important toward the Holocene (Fig.9). The moisture availability influence peaks at 7.5 ka, indicating a widespread shift from energy- to water-limited environments in the early Holocene before shifting back to higher energy-limited influences in the late Holocene. The reconstructions reflect the transitions between energy- and water-limited regimes (in terms of their correlation with the emulations), with a maximum area of water-limited regimes reached at 8.5ka.“

Fig. 5 and Fig. 6 are central but hard to understand; slightly stronger guidance in the captions on how to read them would help non-specialist readers.

A: We changed the captions to:
“Figure 5: Results of different metrics for determining the differences between the simulated and reconstructed tree cover time-series per grid-cells. These are a) Pearson correlation coefficient, b) mean absolute error (MAE), and c) differences in the tree cover variance (Model – Reconstructions). For easier comparability, the results have been grouped in statistical categories by the median value (see Tab. 1). In b) the regions with highest MAE are marked with rectangles Higher correlation and lower MAE indicate better agreement between simulation and reconstruction, while larger variance differences highlight regions where the model fails to capture temporal variability (blue=model overestimates variance, red= model underestimates variance).”

„Figure 6: Difference between simulated tree cover (MPI-ESM) and reconstructed tree cover (REVEALS) across REVEALS tree cover classes. The boxplots show the distribution of tree cover differences (MPI – REVEALS) at any time in relation to the reconstructed REVEALS tree cover class (0–20%, 20–40%, ..., 80–100%). The horizontal line in each box indicates the median difference, while the black dot represents the mean. Positive values indicate that the simulation overestimates tree cover compared to REVEALS, while negative values indicate underestimation.”

Figure 11 is central for interpreting the effect of bias correction, but it is not immediately intuitive that the colours represent percentile-based changes relative to the original simulation rather than absolute agreement. I recommend clarifying this more explicitly in the caption (e.g. blue indicates relative improvement compared to MPI-ESM, not necessarily good agreement) and possibly adding a short explanatory sentence in the main text to guide readers.

A: done: we changed the text in the discussion to:

LL510: “The trained emulator is forced with a bias-corrected climate to test the impact of potential, systematic climate offsets on the tree cover change. The bias-corrected climate and examples of the resulting tree cover emulation in comparison with the original simulation is provided in Fig. B1-B3 in the Appendix B. To evaluate where the bias-corrected emulation enhances model-data agreement (i.e. it shows higher r or lower MAE values or variance differences), we statistically compare the distribution of grid cells with relative improvement (> 25% percentile) or worsenings (< -25% percentile) for the correlation-, MAE-, and variance-groups, using the same classification as in Fig. 5 and Tab.1. Note that a relative improvement in model-data agreement does not necessarily imply a good absolute agreement, as r or MAE values may still indicate substantial discrepancies.”

and we changed the caption to:

“Figure 11: Impact of the bias correction on the three performance metrics (correlation, MAE, and variance). The left panels display spatial distributions of significant changes in these metrics after bias correction. Blue indicates relative improvement compared to MPI-ESM (> 25% percentile), orange relative worsening (< -25% percentile), and grey no change. The right panels summarize these effects by showing the proportion of grid cells with each type of change across the different model-performance categories (mae-, correlation- and variance-groups). Note that a relative improvement (blue) does not necessarily imply good absolute agreement; it only indicates better agreement in the emulation compared to the original MPI-ESM simulation.”

A: Regional time-series examples (Fig. 12) are very effective and could be referenced more explicitly earlier in the text.

R: We agree that the time-series are very helpful to see the differences between the model, emulation and REVEALS, but due to the structure of the paper (first discussing REVEALS- MPI-ESM mismatches and then the improvement/worsening by the emulation) the plot cannot be placed earlier. Therefore, we have not changed it.

References

Dallmeyer, A., Claussen, M., and Brovkin, V.: Harmonising plant functional type distributions for evaluating Earth system models, Climate of the Past, 15, 335–366, https://doi.org/10.5194/cp-15-335-2019, 2019.

Dallmeyer, A., Kleinen, T., Claussen, M., Weitzel, N., Cao, X., and Herzschuh, U.: The deglacial forest conundrum, Nat Commun, 13, 6035, https://doi.org/10.1038/s41467-022-33646-6, 2022.

Dallmeyer, A., Poska, A., Marquer, L., Seim, A., and Gaillard, M.-J.: The challenge of comparing pollen-based quantitative vegetation reconstructions with outputs from vegetation models – a European perspective, Climate of the Past, 19, 1531–1557, https://doi.org/10.5194/cp-19-1531-2023, 2023.

Githumbi, E., Fyfe, R., Gaillard, M.-J., Trondman, A.-K., Mazier, F., Nielsen, A.-B., Poska, A., Sugita, S., Woodbridge, J., Azuara, J., Feurdean, A., Grindean, R., Lebreton, V., Marquer, L., Nebout-Combourieu, N., Stančikaitė, M., Tanţău, I., Tonkov, S., Shumilovskikh, L., and LandClimII data contributors: European pollen-based REVEALS land-cover reconstructions for the Holocene: methodology, mapping and potentials, Earth Syst. Sci. Data, 14, 1581–1619, https://doi.org/10.5194/essd-14-1581-2022, 2022.

Hao, S., Zhang, X., Duan, Y., Gowan, E. J., Zhu, J., Cauquoin, A., Chen, J., Werner, M., and Chen, F.: Model seasonal and proxy spatial biases revealed by assimilated mid-Holocene seasonal temperatures, Science Bulletin, 70, 2014–2022, https://doi.org/10.1016/j.scib.2025.03.039, 2025.

Harrison, S. P., Bartlein, P. J., Izumi, K., Li, G., Annan, J., Hargreaves, J., Braconnot, P., and Kageyama, M.: Evaluation of CMIP5 palaeo-simulations to improve climate projections, Nature Clim Change, 5, 735–743, https://doi.org/10.1038/nclimate2649, 2015.

Kaplan, J. O., Krumhardt, K. M., Gaillard, M.-J., Sugita, S., Trondman, A.-K., Fyfe, R., Marquer, L., Mazier, F., and Nielsen, A. B.: Constraining the Deforestation History of Europe: Evaluation of Historical Land Use Scenarios with Pollen-Based Land Cover Reconstructions, Land, 6, 91, https://doi.org/10.3390/land6040091, 2017.

Li, C., Dallmeyer, A., Ni, J., Chevalier, M., Willeit, M., Andreev, A. A., Cao, X., Schild, L., Heim, B., Wieczorek, M., and Herzschuh, U.: Global biome changes over the last 21 000 years inferred from model–data comparisons, Climate of the Past, 21, 1001–1024, https://doi.org/10.5194/cp-21-1001-2025, 2025.

Li, F., Gaillard, M.-J., Cao, X., Herzschuh, U., Sugita, S., Ni, J., Zhao, Y., An, C., Huang, X., Li, Y., Liu, H., Sun, A., and Yao, Y.: Gridded pollen-based Holocene regional plant cover in temperate and northern subtropical China suitable for climate modelling, Earth System Science Data, 15, 95–112, https://doi.org/10.5194/essd-15-95-2023, 2023.

Liu, Z., Zhu, J., Rosenthal, Y., Zhang, X., Otto-Bliesner, B. L., Timmermann, A., Smith, R. S., Lohmann, G., Zheng, W., and Elison Timm, O.: The Holocene temperature conundrum, Proc. Natl. Acad. Sci. U.S.A., 111, https://doi.org/10.1073/pnas.1407229111, 2014.

Schild, L., Ewald, P., Li, C., Hébert, R., Laepple, T., and Herzschuh, U.: LegacyVegetation: Northern Hemisphere reconstruction of past plant cover and total tree cover from pollen archives of the last 14 kyr, Earth System Science Data, 17, 2007–2033, https://doi.org/10.5194/essd-17-2007-2025, 2025.

Sun, Y., Xu, Q., Gaillard, M.-J., Zhang, S., Li, D., Li, M., Li, Y., Li, X., and Xiao, J.: Pollen-based reconstruction of total land-cover change over the Holocene in the temperate steppe region of China: An attempt to quantify the cover of vegetation and bare ground in the past using a novel approach, CATENA, 214, 106307, https://doi.org/10.1016/j.catena.2022.106307, 2022.

Thompson, A. J., Zhu, J., Poulsen, C. J., Tierney, J. E., and Skinner, C. B.: Northern Hemisphere vegetation change drives a Holocene thermal maximum, Science Advances, 8, eabj6535, https://doi.org/10.1126/sciadv.abj6535, 2022.

Tian, Z. and Jiang, D.: Mid-Holocene ocean and vegetation feedbacks over East Asia, Climate of the Past, 9, 2153–2171, https://doi.org/10.5194/cp-9-2153-2013, 2013.

Zhang, P., Luo, Y., Liu, D., Wang, X., and Wang, T.: Pollen-based reconstruction of spatially-explicit vegetation cover over the Tibetan Plateau since the last deglaciation, Earth System Science Data, 17, 5557–5570, https://doi.org/10.5194/essd-17-5557-2025, 2025.

Zhang, Y., Renssen, H., Seppä, H., and Valdes, P. J.: Holocene temperature evolution in the Northern Hemisphere high latitudes – Model-data comparisons, Quaternary Science Reviews, 173, 101–113, https://doi.org/10.1016/j.quascirev.2017.07.018, 2017.

Zhang, Y., Renssen, H., Seppä, H., and Valdes, P. J.: Holocene temperature trends in the extratropical Northern Hemisphere based on inter-model comparisons, Journal of Quaternary Science, 33, 464–476, https://doi.org/10.1002/jqs.3027, 2018.

Reply to Referee comments on egusphere-2025-6393: „Unravelling the tree cover dynamics over the last 20,000 years on the Northern Hemisphere" by Anne Dallmeyer et al.

Reviewer 2:

Summary: Dallmeyer et al. attempt to unravel the climate drivers of Northern Hemisphere tree cover dynamics over the last 20,000 years using the pollen-based REVEALS reconstruction, a transient MPI-ESM simulation, and a statistical emulator with bias-corrected climate forcing. I commend the authors on the addressing several complex scientific topics, from proxy-model agreement to climate drivers of vegetation dynamics to dynamic vegetation modeled processes, as well as using several independent linear and non-linear statistical measures. They have several interesting findings, including the ability of MPI-ESM to capture broad trends in tree cover and water- vs energy-limited conditions over the last deglaciation and Holocene, but substantial regional data-model mismatches that cannot be resolved with bias correction of climate information, and the dominance of summer temperature as a climate driver in simulated high-latitude forests. Overall, I feel that the scope of the paper may be too large and therefore is difficult to follow as a reader, but with some clarifications and improvements it will be appropriate for the Climate of the Past. I explain some minor and major comments in more detail below.

We thank the reviewer for the positive assessment of our study and for the constructive and helpful comments. We greatly appreciate the time and effort invested in evaluating our work.

Our detailed responses to the reviewer's comments are provided below. For clarity, the reviewer's comments are shown in black, and our responses are given in blue. We have carefully considered all suggestions and have revised the manuscript accordingly to improve its clarity and quality.

We agree that the manuscript covers several interconnected topics. To improve readability, we revised parts of the manuscript to clarify the structure of the Results and Discussion section. In particular, we added a short introductory paragraph at the beginning of the Results and Discussion section to explicitly outline its structure and the guiding research questions, thereby better guiding the reader through the different analyses. We believe these revisions improve the clarity and logical flow of the manuscript while maintaining its overall scope. We have added:

LL227: "We structure our analysis around a sequence of interconnected questions that progressively link model evaluation with process understanding. We begin by comparing simulated and reconstructed tree cover dynamics from the Last Glacial Maximum to the present (Sect. 3.1), establishing the overall level of agreement between MPI-ESM and REVEALS. Building on this, we quantify and characterize the spatio-temporal differences between both datasets (Sect. 3.2), thereby identifying where and when discrepancies occur.

In a next step, we investigate the underlying mechanisms by analysing the main drivers of tree cover change (Sect. 3.3). This process-based perspective is further refined by explicitly contrasting regions with high and low model–data agreement to assess which factors constrain tree cover dynamics in the model (Sect. 3.4). Finally, we address the question whether a bias-corrected climate can reduce the identified model–data differences (Sect. 3.5), thereby assessing the influence of systematic climate biases on simulated tree cover dynamics."

Minor Comments:

Line 163: grid should be capitalized → done

Line 233 and Figure 8: The panels are not labeled with letters but are referenced in the text (Fig. 8d)

A: We added panel labels to the figure.

Major Comments:

R: This study has a high number of acromyns and shortened words (e.g., BiasCorr_Variance, F_CO2, r-group-1) that make it very difficult to know what exactly is being discussed. I recommend lengthening and clarifying the group names to make them more self-explanatory so that readers do not need to study the main text and/or figure captions before looking at each plot. Two examples of this: (1) Table 1: why not use the "Group" as "group_name" so that the meaning of the groups is more clear in the text? Naming a group "Positive Correlation Group" rather than "r_group 1" would be easier to understand. (2) Figure C1: why do you use 1, 2, and 3 as x-axis labels rather than writing out Positive, None, Negative Correlations? This adds an unnecessary layer of complexity to an already complex paper.

A: Thank you for this helpful comment. We agree that the original naming (e.g., "r_group 1" and numeric axis labels) was not sufficiently intuitive and could make interpretation more difficult for the reader.

In response, we have revised the group names throughout the manuscript and now use more descriptive terms (e.g., "positive-correlation-group," "high-MAE-group", ...) in both the text and tables. We have also clarified and standardized abbreviations such as *F_* and *R_* in the main text to improve readability and define the coefficients and variables in Tab. C1.

At the same time, we have changed the group names etc., but retained abbreviated labels within the figures themselves to avoid overcrowding and maintain visual clarity. We believe this provides a good balance between readability in the manuscript and clarity in the graphical presentation. To implement this, we have changed the following figures: Fig. 5, Fig. 10, Fig. 11 and Fig. 12.

R: A video animation of the forest cover through time would be incredibly powerful and helpful for following the description of results, a great example of this is Shafer et al. (2021).

A: This is a valuable suggestion. We provide a video illustrating the vegetation pattern for all time slices (https://doi.org/doi:10.17617/3.O2GGHG) and add the link to the caption of Fig.4. In addition, we refer here to another video showing the biome distribution based on the simulation used in this study alongside reconstructions (https://hdl.handle.net/21.11116/0000-000B-2420-8).

R: Lines 327-329: How did you determine this list of climate drivers to consider? I am wondering why other variables, such as incident solar insolation and mean annual temperature, were not included.

A: The selection of variables is motivated by their known importance for vegetation productivity and bioclimatic thresholds implemented in the model (Tc, Tw). We have not performed a statistically rigorous variable performance but did test other climate drivers such as annual mean temperature, insolation, and growing degree days, but did not find an added value from including other drivers. Including too many drivers, that are partly dependent on each other and those don't provide much independent information, also increases the risk of overfitting and inferring non-causal ecological relationships. Therefore, we include climate variables that represent the important climatic growth limitations from growing season warmth, winter cold, and moisture availability, in addition to CO2 for representing plant-physiological CO2 influences.

We added to the manuscript:

L186: "The selection of these variables is motivated by their known importance for vegetation productivity and the implementation of bioclimatic thresholds for Tw and Tc in JSBACH."

R: Line 453-454: What climate variables/regions changed the most dramatically with bias correction? It would be very useful to see a couple time slice plots where the bias-corrected climate makes a large difference in simulated tree cover. For example, on lines 588-591, you mention that the model is too cold in the northern high-latitudes. This would be helpful to understand your interpretation of impacts of bias correction on simulated tree cover.

A: Thank you for this helpful suggestion. We agree that spatial snapshots of the emulator output and climate forcing would be useful to better illustrate the impact of bias correction.

However, because we apply a delta method, the temporal evolution of anomalies is preserved: the difference between the bias-corrected and simulated Twarm and Tcold remains constant over time, and precipitation mainly changes in amplitude rather than spatial pattern. We therefore focus on time series of Northern Hemisphere means to represent temporal changes, complemented by maps at selected time slices.

Specifically, we decided to show one climate snapshot at the time of maximum change (~12 ka BP), along with tree cover and its differences for three representative periods. We added the following figures to the Appendix B:

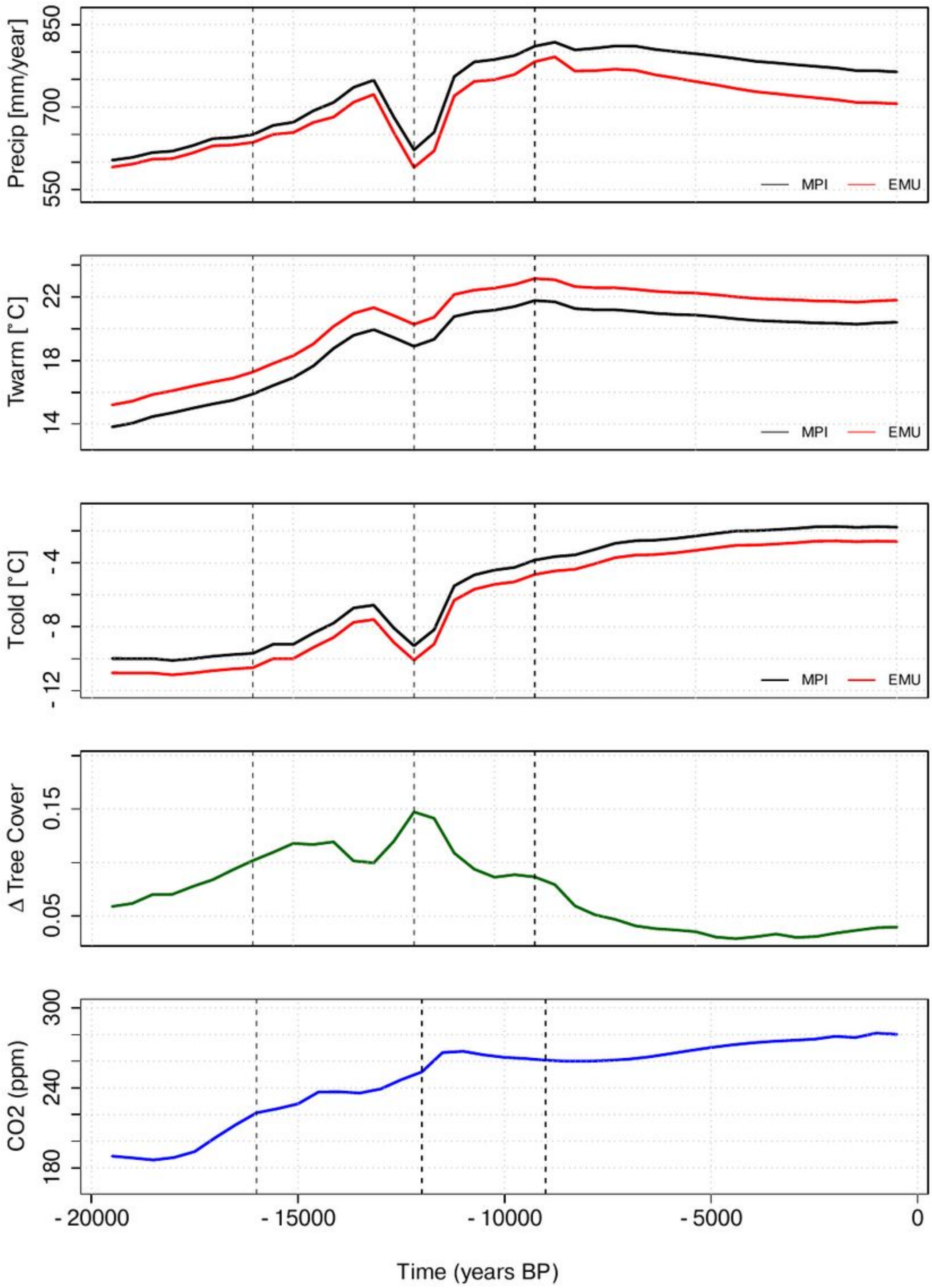

Fig. B1: Time series of Northern Hemisphere mean precipitation (precip), temperature of the warmest (Twarm) and coldest month (Tcold) for the original (MPI) and bias-corrected (EMU) climate, together with the corresponding difference in mean tree cover between the bias-corrected emulation and the original simulation. Shown is also the global atmospheric CO2 concentration that is used as predictor in the emulation. The dotted lines mark the three time-slices for which maps of tree cover are provided in Fig. B2

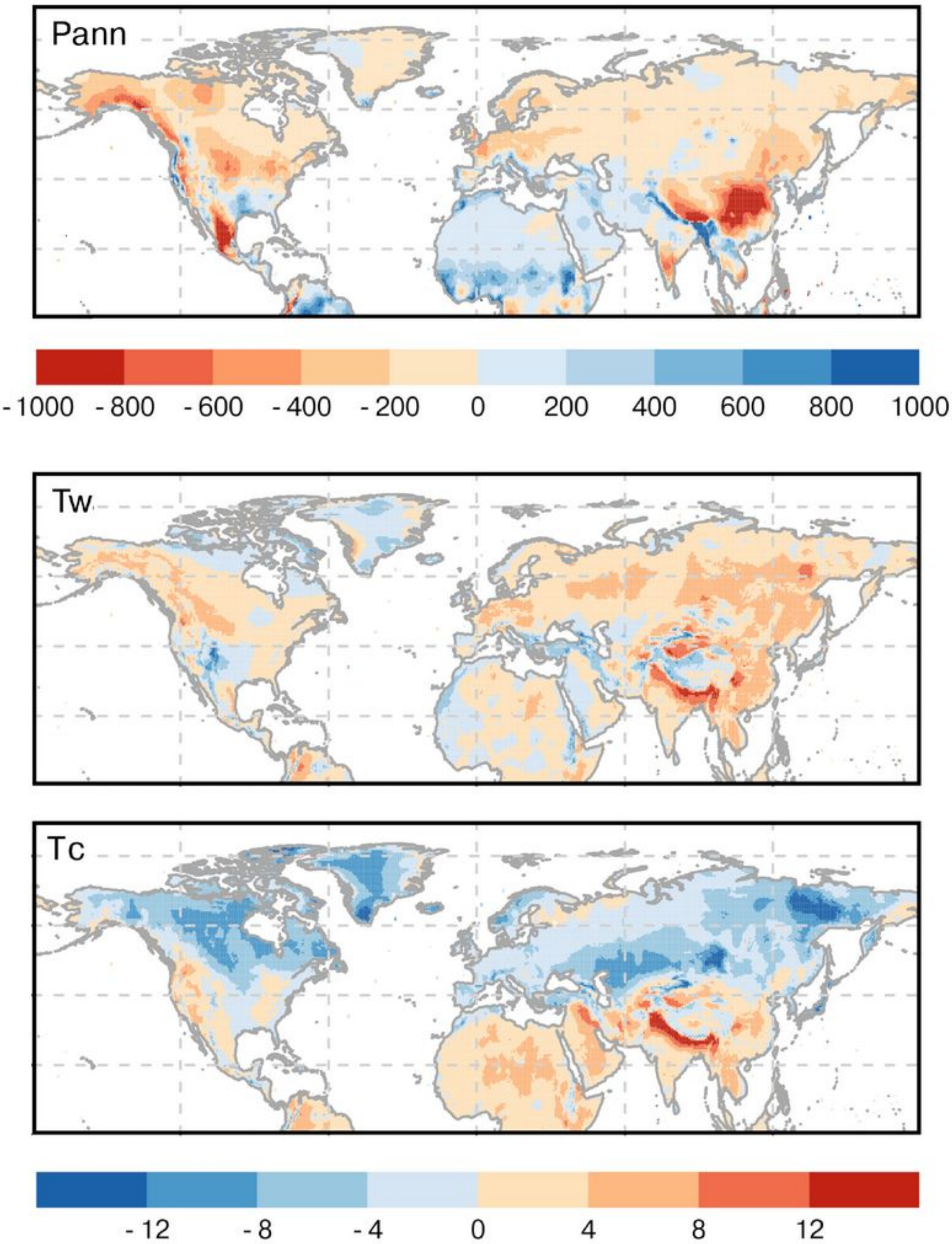

Fig. B2: Difference in annual mean precipitation (Pann) [mm/year] and in the temperature of the warmest (Tw) and coldest (Tc) month [K] at 12,000 years BP between the bias-corrected and the original simulation. The difference was calculated after remapping the simulation onto a 0.5° grid.

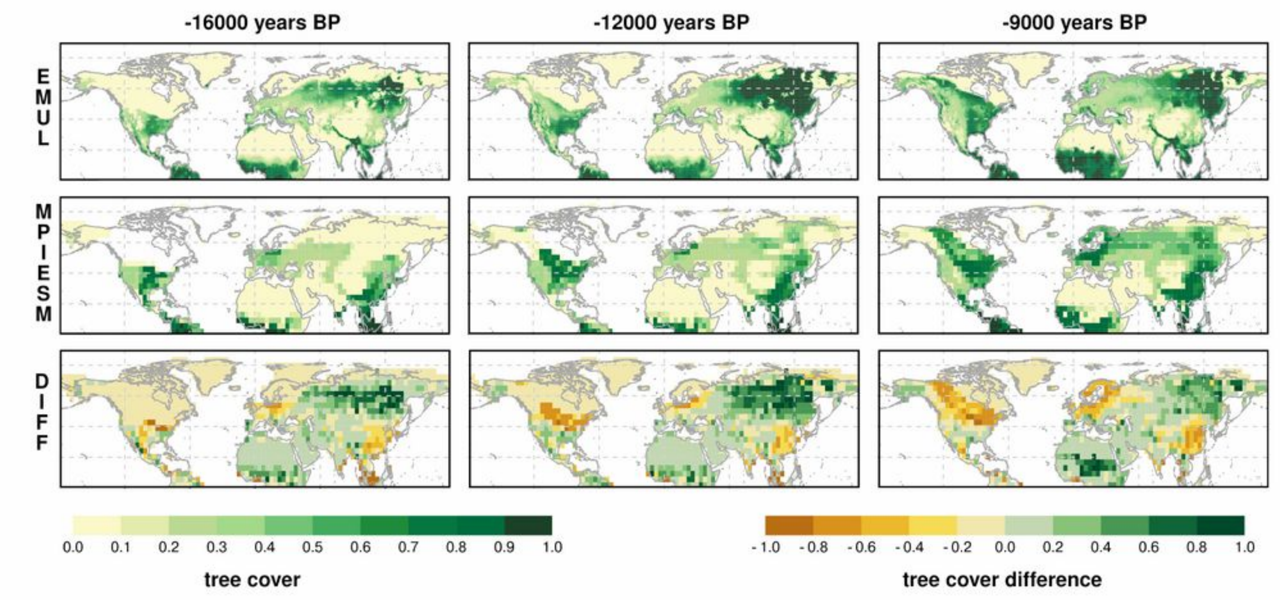

Fig. B3: Tree cover fraction at different time steps according to the emulation with bias-corrected climate and the MPI-ESM simulation. The bottom panel also shows the difference between the two; for this, the emulation has been remapped to a T31 grid.

We add the following text to the Appendix B (LL760): „Time series of Northern Hemisphere mean precipitation, warm- and cold month temperature for both the original and bias-corrected climate, together with the corresponding difference in mean tree cover between the bias-corrected emulation and the original simulation are shown in Fig. B1 in the Appendix B. In addition, spatial maps of the climate variables at the time of maximum change (~12 ka BP), as well as maps of emulated tree cover, simulated tree cover, and their difference for three representative time slices, are provided in Fig B2 and B3 in the Appendix B.“

And to Section 3.5 (LL511): “The bias-corrected climate and examples of the resulting tree cover emulation in comparison with the original simulation is provided in Fig. B1-B3 in the Appendix B.“

R: Lines 582-583: I’m not entirely convinced that the tree cover response is too sensitive to CO2-level. This implies that tree cover is responding very strongly to the physiological impacts of CO2 concentrations, which should play some role, but I am not sure that you have included all of the potential climate predictors that capture the indirect, temperature-related effects of CO2. You have included two temperature-related predictors, Tw and Tc. Do these strongly correlate with CO2 level, or does mean annual temperature more closely vary with CO2 level? Can you clarify the implications of “the sensitivity of the tree cover response to CO2-changes seem to be too strong in this model version” on the way that photosynthesis/stomatal conductance is represented in MPI-ESM and JSBACH?

A: We agree that the strong tree cover response to CO_2 does not necessarily imply an unrealistically strong direct physiological effect alone. In MPI-ESM, vegetation responds to CO_2 through coupled physiological processes and indirect climate feedbacks, and the signal emerges across multiple intermediate scales between leaf-level photosynthesis in JSBACH and biome-scale vegetation dynamics. However, we think that the temperature-related effects of CO_2 are mostly accounted for by the included temperature predictors (Tw and Tc). We additionally tested annual mean temperature in the GAM and correlation analyses, and it didn't provide any added value compared to including Tw and Tc.

The strong CO_2 sensitivity is consistent with CMIP6 analyses (e.g., Arora et al., 2020), which show that MPI-ESM exhibits a relatively high CO_2 fertilisation effect compared to the multi-model mean. Furthermore, Kleinen et al. (2023) suggest that the high CO_2 sensitivity in MPI-ESM may contribute to model–data mismatches in methane levels since the last deglaciation.

Structural aspects of the photosynthesis formulation may contribute to the high sensitivity. The current JSBACH implementation has a comparatively weak temperature constraint relative to other models. As shown by Rogers et al. (2017), leaf-level productivity increases substantially from preindustrial to elevated CO_2 levels (e.g., 380 to 550 ppm), indicating a strong physiological CO_2 response already at canopy scale.

In addition, limited water stress representation and the absence of nutrient limitations may further enhance the simulated CO_2 fertilisation effect.

To support our statement in the revised version, we now point to previous studies that have documented the high CO_2 sensitivity in MPI-ESM and related JSBACH configurations.

We have added in the revised manuscript (LL506): „This high sensitivity to CO_2 in MPI-ESM has also been observed in other studies (Arora et al., 2020; Kleinen et al., 2023a; Rogers et al., 2017).

R: Section 3.3: The emulator approach to disentangle the importance of different climate drivers is very interesting. In Figure 7, you visualize the most important climate driver in each grid cell, but I'm wondering how often there is a mixture of variables that have substantial importance (e.g., let's say in N America, P and Tw are always the highest correlations so in many grid cells that are light blue or orange there is only a small difference their relative importance and which one is "most important") and that information is missing from this figure. Would it be possible to add this information by modifying the shade of the colors? For example, if P is the most important predictor in one grid cell and far outcompetes other predictors, it is light blue, but if it is only slightly more important than other predictors, it is closer to white.

Thank you for this helpful suggestion. We agree that Figure 7, in its current form, does not convey how strongly the leading predictor outcompetes the other variables, and therefore may obscure regions where multiple drivers have similar importance. To address this, we will revise the figure by marking the grid cells with black dots where the difference in correlation coefficients between the leading predictor and the second most important predictor exceeds 0.1, which corresponds to the median of all differences. This will help highlight regions with significant differences more clearly.

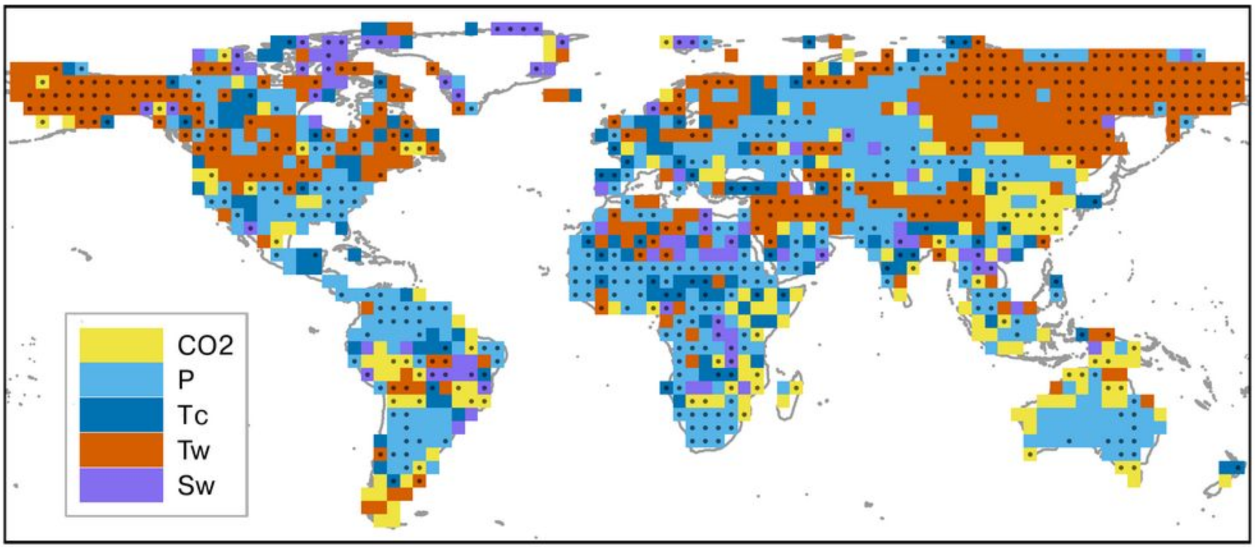

Figure 7: Spatial distribution of the most influential climatic variables determining vegetation dynamics across different regions of the world. Each grid cell is color-coded according to the dominant driver of vegetation change in the emulations, which is here based on the emulation with single variable forcing that show the highest Pearson correlation to the simulated tree cover. The variables considered are atmospheric CO_2 concentration (yellow), annual precipitation (P, light blue), temperature of the coldest month (Tc, dark blue), temperature of the warmest month (Tw, orange) and annual mean soil moisture (Sw,purple).Grid cells marked with black dots indicate regions where the difference in correlation coefficients between the leading predictor and the second most influential predictor exceeds 0.1 (the median difference), highlighting areas with a comparatively strong dominance of a single driver.

References

Arora, V. K., Katavouta, A., Williams, R. G., Jones, C. D., Brovkin, V., Friedlingstein, P., Schwinger, J., Bopp, L., Boucher, O., Cadule, P., Chamberlain, M. A., Christian, J. R., Delire, C., Fisher, R. A., Hajima, T., Ilyina, T., Joetzjer, E., Kawamiya, M., Koven, C. D., Krasting, J. P., Law, R. M., Lawrence, D. M., Lenton, A., Lindsay, K., Pongratz, J., Raddatz, T., Séférian, R., Tachiiri, K., Tjiputra, J. F., Wiltshire, A., Wu, T., and Ziehn, T.: Carbon–concentration and carbon–climate feedbacks in CMIP6 models and their comparison to CMIP5 models, Biogeosciences, 17, 4173–4222, https://doi.org/10.5194/bg-17-4173-2020, 2020.

Kleinen, T., Gromov, S., Steil, B., and Brovkin, V.: Atmospheric methane since the last glacial maximum was driven by wetland sources, Climate of the Past, 19, 1081–1099, https://doi.org/10.5194/cp-19-1081-2023, 2023.

Rogers, A., Medlyn, B. E., Dukes, J. S., Bonan, G., von Caemmerer, S., Dietze, M. C., Kattge, J., Leakey, A. D. B., Mercado, L. M., Niinemets, Ü., Prentice, I. C., Serbin, S. P., Sitch, S., Way, D. A., and Zaehle, S.: A roadmap for improving the representation of photosynthesis in Earth system models, New Phytologist, 213, 22–42, https://doi.org/10.1111/nph.14283, 2017.