Global biome changes over the last 21,000 years inferred from model-data comparisons

Response to comments of editor

Reviewer comment: (1) *Further clarification on the role of human land use and fire in model-data mismatches would enhance the discussion.*

Response: Thank you for your comment. The simulation we used included fire but not land use, which was stated in the text, so we have clarified the role of land use in the data-model comparison in the revised text.

Text: Line 169-174: "The dynamic vegetation in all models is represented by different sets of plant functional types (PFTs) that can coexist in the grid-cells. The occurrence of each PFT is constrained by fixed temperature thresholds, and the dynamics of PFT cover fraction depends for instance on the moisture availability and plant requirements. Disturbances such as fire, which are already coupled in the dynamic vegetation modules, regularly reduce the coverage of tree and shrub PFTs while promoting the expansion of herbaceous PFTs (Burton et al., 2019; Reick et al., 2021; Dallmeyer et al., 2022). Land use is not included in any of these simulations.**"**

Revised text: (in red)

Line 464-481: "Different estimates of non-forest megabiomes in relatively semi-arid zones, such as North Africa and the Mediterranean, have contribute to moderate but increasing data-model deviations since the early deglaciation (Fig. 4e). As shown in Fig. 3, with the transition from the glacial to the Holocene, the Mediterranean-Black Sea-Caspian Corridor (EU2) and the Mediterranean coast of northern Africa have gradually been dominated by temperate forests (TEFO) in the reconstructions, rather than grasslands and dry shrublands (STEP) in the simulations. Since the reconstructions better reproduces the region's modern potential natural vegetation than the simulation (Table 3), we infer that the simulations likely underestimated the cover fraction of woody PFTs in the simulations throughout the Holocene. Given that anthropogenic disturbances (e.g., land use and deforestation) did not promote large-scale forest degradation in this region (cf. Sect. 3.2), this underrepresentation could be attributed to the systematic model biases of hotter summers and drier winters (García-Herrera and Barriopedro, 2018; Fig. A3a–b). In addition, data-model deviations in the Sahara (AF1) are primarily observed during the Holocene, resulting from a mismatch between simulated deserts (DESE) and reconstructed savanna (SAVA). In the simulations, the weakening of the North African monsoon system led to desert expansion in response to seasonal insolation changes, a pattern supported by both proxy-based reconstructions (deMenocal et al., 2000; Shanahan et al., 2015) and climate simulations (Dallmeyer et al., 2021). However, in our reconstructions, the overrepresentation of woody taxa (e.g., *Acacia* and Arecaceae) resulted in the classification of some desert regions as savanna and dry woodlands (SAVA), potentially contributing to the increasing data-model deviations in the Sahara during the Holocene."

Reviewer comment: (2) *The authors should ensure a precise explanation of biome interpolation in ESM grids.*

Response: Thank you for your comment and we have revised the text.

Revised text: (in red)

Line 244-250: "We aggregated the records into regular longitude-latitude grid-cells of size 3.75°x3.75° to reduce the sampling bias from the non-uniform spatial distribution of records and to facilitate a more direct model-data comparison. At each timeslice, the reconstructed or simulated megabiome assigned to a grid-cell was determined based on the most frequently occurring megabiome among the available records in that grid-cell. When multiple megabiomes had the same highest frequency, we applied the same criterion used in pollen-based reconstructions, prioritizing the highest-frequency megabiome with the fewest PFTs and taxa. Similarly, the data-model EMDs for each grid-cell were derived as the median EMDs of the available records within that grid-cell."

Reviewer comment: (3) *Minor technical clarifications, such as references for newly included pollen records, should be incorporated.*

Response: We newly added the description of the archived dataset in the revised text.

Revised text: (in red)

Line 97-107: "The LegacyPollen 2.0 dataset is archived in the PANGAEA repository (https://doi.org/10.1594/PANGAEA.965907; Li et al., 2025) and is open-access. It follows the framework of LegacyPollen 1.0 dataset (Herzschuh et al., 2021, 2022), providing pollen count and pollen percentage data per continent, a taxa harmonization master table, and site metadata (such as data sources, Dataset ID, site name, location, archive type, site description, and references). To enhance data traceability and ensure high-quality standards, we have newly incorporated the Neotoma digital object identifier (DOI) into the metadata for Neotoma-derived records, allowing direct linkage to the living Neotoma database and reducing the risk of data staleness. These DOIs were generated with the *doi*() function from the package *neotoma2* (version 1.0.3; Socorro and Goring, 2024) in the R software environment (version 4.4.1; R Core Team, 2023). Additionally, we also newly added the PANGAEA Event (PANGAEA dataset identifier) for each record to ensure that our dataset meets PANGAEA's high standards for quality, usability, and compliance."