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

Response to comments of Referee #1 Guiot Joel

1. General comments

Reviewer comment: (1) *Authors present a new global dataset of 3691 pollen records that are transformed into biomes using the well-established method of Prentice et al (1996). The dataset covers the whole globe with some gaps in Africa, southern Asia and Australia, where the data are not available.*

Response: We acknowledge the limited availability of data in certain regions and have clarified this in the text.

Text: "While there are geographical gaps in pollen record coverage, particularly in the Southern Hemisphere, the dataset LegacyPollen 2.0 covers the world's main vegetation and climate zones."

Reviewer comment: (2) *They work at the level of 8 mega biomes which represent raw pattern of global vegetation. Indeed, it is difficult to work with finer biomes at a global scale (even if Prentice at al 2000 did for two time slices). The biomes reconstructed from 21 ka BP to present at 500 years' time steps are compared with equivalent biomes simulated by three models (and two simulations each).*

Response: Thank you for your comment. We have revised the text to clarify the classification of megabiomes and its application in data-model comparison as follows:

Revised text: (in red)

"The PFTs were then assigned to megabiomes, representing the raw pattern of global vegetation rather than the finer biomemegabiome categories commonly used in standard biomization studies (Dallmeyer et al., 2019). These megabiomes include), namely tropical forest (TRFO), warm-temperate (subtropical) forest (WTFO), temperate forest (TEFO), boreal forest (BOFO), (warm) savanna and dry woodland (SAVA), grassland and dry shrubland (STEP), (warm) desert (DESE), tundra and polar desert (TUND). These categories wereare also applied to biomize Earth System Model results, which generally use different types and numbers of PFTs to represent global vegetation, enablingallowing for direct data-model comparisons and evaluations (Dallmeyer et al., 2019)."

Reviewer comment: (3) *The main problem of the paper can be easily detected in the abstract (which reflect correctly the full paper). The paper is technical and too descriptive.*

Response: Thank you for your comment. We have revised the abstract to better emphasize the key findings and broader implications while reducing technical descriptions.

Revised text:

"We present a global megabiome reconstruction for 43 timeslices at 500-year intervals throughout the last 21,000 years based on an updated and thus currently most extensive global taxonomically and temporally standardized fossil pollen dataset of 3,455 records. The evaluation with modern potential natural vegetation distributions yields an agreement of ~80%, suggesting a high degree of reliability of the pollen-based megabiome reconstruction. With its high temporal and spatial resolution, this reconstruction is a robust dataset for various applications, such as the evaluation of paleo-simulations from Earth System Models (ESMs).

We compare the reconstruction with an ensemble of six different biomized simulations based on transient vegetation simulations performed by ESMs. The global spatiotemporal patterns of megabiomes estimated by the simulation ensemble and reconstructions are generally consistent. That is, there has been a global shift from open glacial non-forest megabiomes to Holocene forest megabiomes since the Last Glacial Maximum (LGM), in line with the general climate warming trend and continental ice-sheet retreat. The shift to a global megabiome distribution generally similar to today's took place during the early Holocene, while enhanced anthropogenic disturbances since the Late Holocene have not altered broad-scale megabiome patterns. However, certain data-model deviations are evident in specific regions and periods, which could be attributed to systematic climate biases in ESMs or biases in the pollen-based biomization method. For example, at a global scale over the last 21,000 years, the largest deviations between reconstructions and simulation ensembles are observed during the LGM and early deglaciation periods. These discrepancies are probably attributed to systematic summer cold biases in the ESMs that lead to an overestimation of tundra in periglacial regions. Additionally, steppes are overrepresented on the Tibetan Plateau, because steppes share dominant characteristic taxa with tundra and are preferential allocated due to fewer plant functional types (PFTs) compared to tundra. Moderate deviations during the Holocene are primarily driven by differing estimates of non-forest megabiomes in the Mediterranean and North Africa, with increasing discrepancies over time. These deviations may result from the underestimation of woody PFT cover in simulations due to systematic biases, such as overly warm summers with dry winters in the Mediterranean, and the overrepresentation of woody taxa in reconstructions, misclassifying deserts as savanna in North Africa."

Reviewer comment: (4) *Biomes are reconstructed, simulated with ESM and compared, some biases of the ESM simulations are pointed out and that is all. There are no general messages on the evolution of the vegetation through the 21 millennia.*

Response: We have described the global vegetation evolution over the past 21,000 years in section 3.1 as follows:

Text: "We present a global assessment of megabiome dynamics and distributions derived from pollen-based reconstructions and ESM-based simulations over the last 21,000 years, with a temporal resolution of 500 years.

Overall, there has been a global shift from open glacial non-forest megabiomes to Holocene forest megabiomes since the LGM (Fig. 2), in line with the general climate warming trend and continental ice-sheet retreat (Fig. 3):

LGM (represented by the timeslice 21 cal. ka BP): TUND and BOFO dominate the high latitudes and periglacial areas (similar to Prentice et al., 2000 and Davis et al., 2024), whereas the relatively warm forest megabiomes (e.g., WTFO and TEFO) are distributed at lower latitudes than present-day, in response to cold and dry climates. [...]"

Reviewer comment: (5) *The ESM biases are not discussed by analysing which climatic variables are responsible of them.*

Response: Thank you for your comment. We performed further analyses to assign the ESM biases to certain climatic variables. Due to the lack of pollen-independent climate reconstructions at broad spatial and long temporal scales, we cannot determine when and why models exhibit climate biases. Consequently, we limited our comparison to modern observed climate data in the revised text.

New text in Method:

"Modern observational climate data provide a crucial foundation for the assessment of climate simulations. The Climatic Research Unit gridded Time Series (CRU TS hereafter), version 4.08, is a widely used modern observational climate dataset covering all land domains of the world except Antarctica (spatial resolution: ~0.5°x0.5° on a Gaussian grid; Harris et al., 2020). The CRU TS dataset is interpolated from extensive networks of weather station observations and provides monthly temperature and precipitation data from 1901-2023 C.E. However, the early record of this dataset may have high uncertainty due to the sparse observation networks (Duan et al., 2024), and the late record is strongly influenced by anthropogenic CO_2 increases (Cheng et al., 2022). We, therefore, selected monthly climatological means from 1931-1970 to generate more biologically meaningful bioclimatic variables for evaluating climate simulations at 0 cal. ka BP (O'Donnell and Ignizio, 2012; Supplementary Data 7). These bioclimatic variables represent extreme or limiting environmental factors, namely, mean temperature of warmest quarter (T_{warm}), mean temperature of coldest quarter (T_{cold}), precipitation of warmest quarter (P_{warm}), and precipitation of coldest quarter (P_{cold}). Temperature is given in degrees Celsius (°C), precipitation in millimeters (mm), and a quarter is a period of three consecutive months (1/4 of the year)."

Revised text in Results and Discussion:

"The agreement between modern potential megabiomes and simulated megabiomes at timeslice 0 cal. ka BP is higher for the ESM-representative megabiome (cf. Sect. 2.3) than for individual ESM-based simulation (64.1% vs. 20.0–60.2%; Table 3). As a result, the ESM-representative megabiome depicts more reliable patterns of megabiome dynamics and distribution than individual simulations, with higher agreement especially in Alaska, the Iberian Peninsula, the Alps, the Atlantic Coastal Plain of North America, and the southeastern United States (Fig. 1 and Fig. A2). However, there are still certain regions with low agreement, probably due to climatic biases. These include nearly all highlands (such as the central-southern Rockies, the central Andes, and the Tibetan

for which an overestimation of the temperature can be expected in the models due to a much lower orography than in reality caused by the smoothing in the coarse spatial resolution (3.75°x3.75° and 5°x5°) of the model grids (Fig. A3a–b). All models simulate non-forest megabiomes instead of forest in the Mediterranean region, which can be attributed to the models simulating a climate that is too seasonally dry, with, for example, too-warm summers and too-dry winters (Fig. A3a, d). The TRACE-21K simulation as well as the MPI-ESM simulations fail to reproduce the boreal forest (BOFO) in Alaska, which is then also reflected in the ESM-representative megabiomes. This failure is likely due to the simulated climate being too cold in this region, preventing the establishment of boreal forests under modeled conditions (Fig. A3a, d). Similar to the reconstructions, the transition zones between temperate forest (TEFO) and non-forest megabiomes, such as the East Asian summer monsoon margin, are regions with lower simulated megabiome agreement to the modern potential megabiome distribution. In North Africa, the models also tend to underestimate the northern extension of the grassland and dry shrubland (STEP) and incorrectly assign (warm) savanna and dry woodland (SAVA) records to tropical forest (TRFO). This is related to the biomization procedure for the model results that only relies on simulated vegetation cover fractions and simulated climate, whereas savannas are additionally determined by other ecological processes such as fire intensity and frequency (Dallmeyer et al., 2019) or grazing (van Langevelde et al., 2019)."
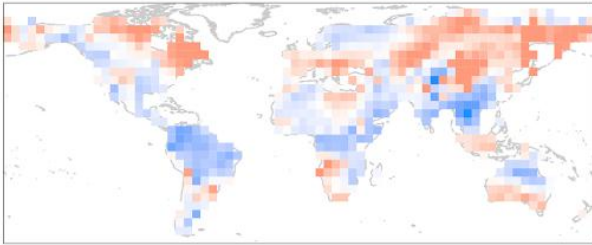
New figure (Appendix A3):

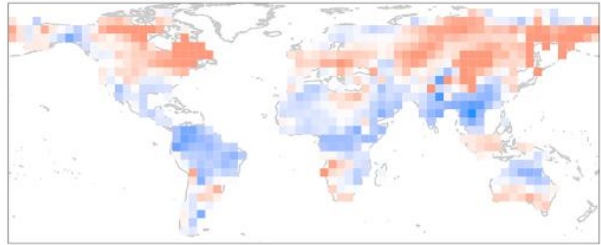**(a) T_Warm: MPI-ESM_GLAC1D - Observation**
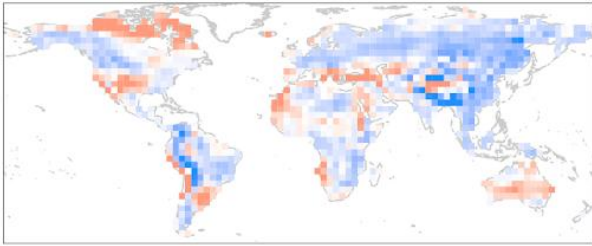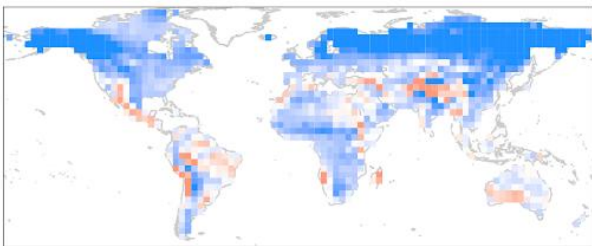
T_Warm: MPI-ESM_ICE6G - Observation
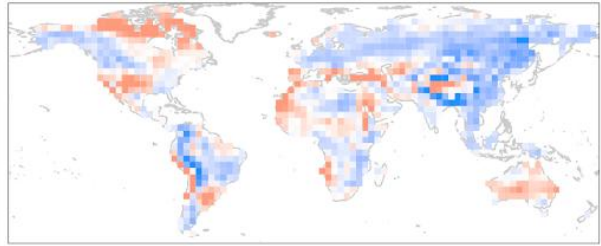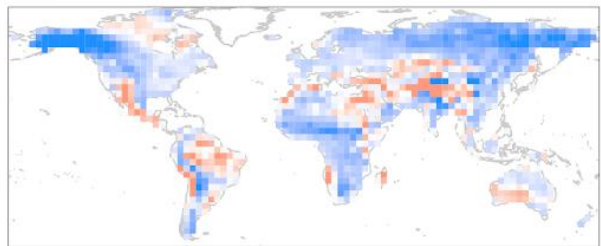
T_Warm: CLIMBER-X_GLAC1D - Observation

T_Warm: CLIMBER-X_ICE6G - Observation

T_Warm: TRACE-21K-I_ICE5G - Observation

T_Warm: TRACE-21K-II_ICE5G - Observation

BIO 10

-8   -4   0   4 [°C]

(b) T_Cold: MPI-ESM_GLAC1D - Observation
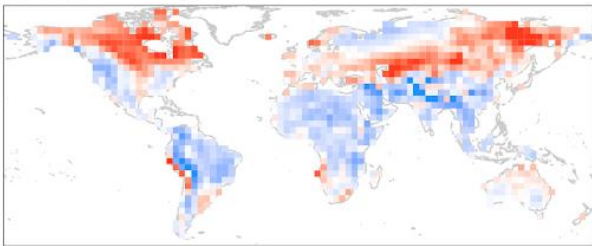
T_Cold: MPI-ESM_ICE6G - Observation

T_Cold: CLIMBER-X_GLAC1D - Observation

T_Cold: CLIMBER-X_ICE6G - Observation

T_Cold: TRACE-21K-I_ICE5G - Observation

T_Cold: TRACE-21K-II_ICE5G - Observation

BIO 11

-8   -4   0   4   8 [°C]

5

(c) P_{Warm}: MPI-ESM_GLAC1D - Observation

P_{Warm}: MPI-ESM_ICE6G - Observation

P_{Warm}: CLIMBER-X_GLAC1D - Observation

P_{Warm}: CLIMBER-X_ICE6G - Observation

P_{Warm}: TRACE-21K-I_ICE5G - Observation
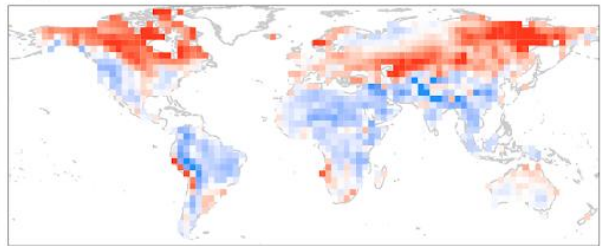
P_{Warm}: TRACE-21K-II_ICE5G - Observation

BIO 18
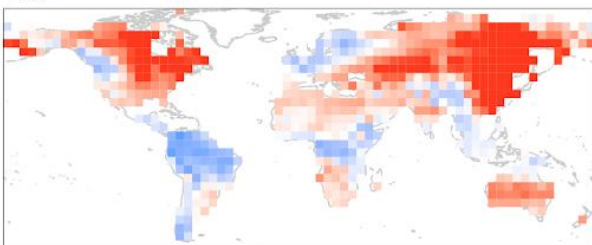-200   -100   0   100   200   [mm]

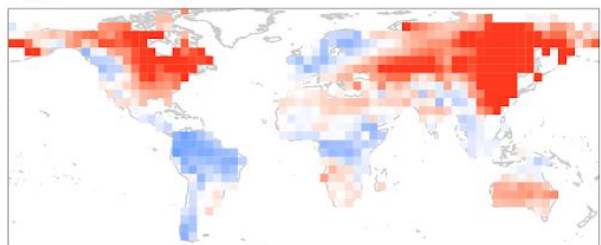**(d) P_{Cold}: MPI-ESM_GLAC1D - Observation**

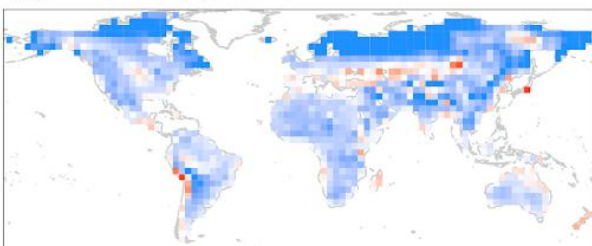P_{Cold}: MPI-ESM_ICE6G - Observation

P_{Cold}: CLIMBER-X_GLAC1D - Observation

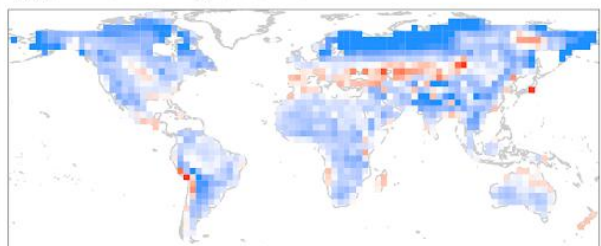P_{Cold}: CLIMBER-X_ICE6G - Observation
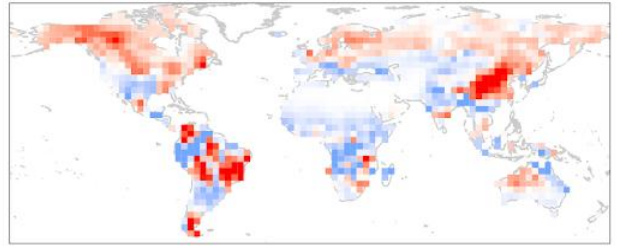
P_{Cold}: TRACE-21K-I_ICE5G - Observation

P_{Cold}: TRACE-21K-II_ICE5G - Observation
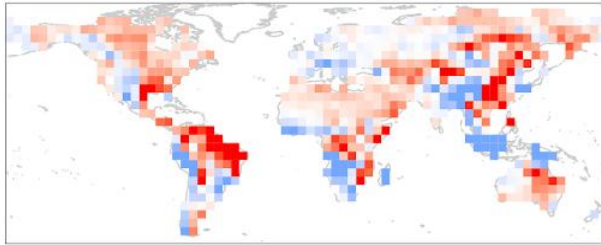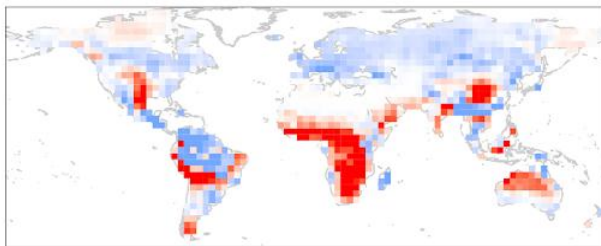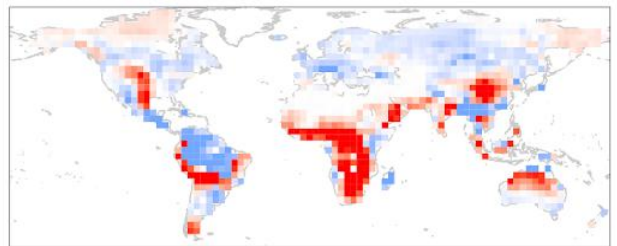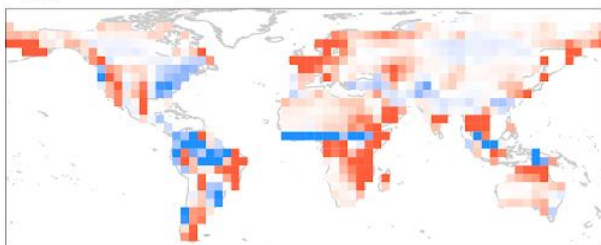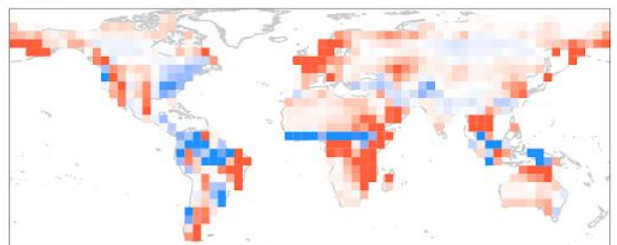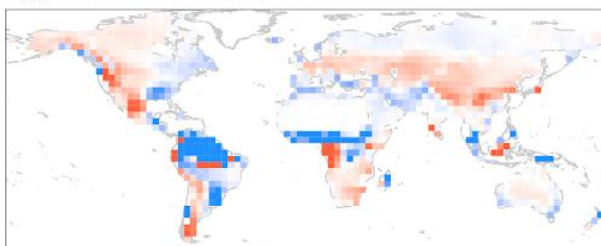
BIO 19
-200   -100   0   100   [mm]

6

Reviewer comment: (6) *Why mediterranean biomes of the Holocene are simulated as steppes whereas data indicate TEDE?*

Response: All models consistently simulate steppe (STEP) rather than temperate forest (TEDE) in the Mediterranean region during the Holocene, likely due to systematic biases in the simulated climate, as described in the revised text.

Revised text:

"The Mediterranean region has warm to hot dry summers and mild wet winters. Modeling studies report systematic model biases of too-warm summers and too-dry winters in this region (García-Herrera and Barriopedro, 2018). A comparison with modern data shows similar climate biases in the simulations which may indicate similar systematic biases in the past. This would explain the underrepresentation of the cover fraction of woody PFTs in the simulations (Fig. A3a, d)."

Reviewer comment: (7) *There are also biases in the reconstructed biomes. As an example, the glacial mediterranean biomes are sometimes reconstructed as TEDO while previous reconstructions (Elenga et al, 2000; Prentice et al, 2000) reconstructed STEP.*

Response: Thank you for your comment. We have added more comparisons in the revised text to clarify the differences between our reconstructions and previous studies.

Revised text: (in red)

"LGM (represented by the timeslice 21 cal. ka BP): TUND and BOFO dominate the high latitudes and periglacial areas (similar to Prentice et al., 2000), whereas the relatively warm forest megabiomes (e.g., WTFO and TEFO) are distributed at lower latitudes than present-day, in response to cold and dry climates (Nolan et al., 2018). However, the ESM-representative megabiome (simulations hereafter in this Sect.) reveals more non-forest megabiomes (such as TUND and STEP) in periglacial areas of North America (e.g., Alaska and the Rocky Mountains) and northern Asia (e.g., northeastern Siberia), as well as in the Mediterranean regions, as compared to the reconstructions. Although previous pollen-based biomization studies with different biomization schemes have reported ESM-like results (such as Binney et al., 2017 and Cao et al., 2019 in periglacial areas; Elenga et al., 2000 and Prentice et al., 2000 in the Mediterranean regions), assessments of modern megabiome distributions

suggest that these studies overestimated the occurrence of non-forest megabiomes in these regions. A recent pollen-based forest cover reconstruction by Davis et al. (2024) indicates more forest than previously suggested by biome reconstructions in these regions during the LGM, which aligns with our results. Furthermore, STEP occurred in central Asia in the reconstructions rather than TUND in the simulations, and TRFO and SAVA appeared in tropical South America and Africa in the reconstructions rather than WTFO in the simulations."

Reviewer comment: (8) *I have the feeling that for the regions I know better, this paper does a poorer reconstruction than previous attempts. If not, this should be argued.*

Response: Thank you for your comment. We have added regional comparisons in the revised text to further clarify the reliability of our reconstruction.

Revised text:

"We consider global-scale, pollen-based megabiome reconstructions reliable because record-by-record comparisons of reconstructed megabiomes at timeslice 0 cal. ka BP from 2,232 available records with modern potential megabiomes indicate an 80.2% agreement (Table 3). This consistency exceeds that reported in previous large-scale biomization studies validated against modern biome distributions, such as the 53% agreement in Arctic high-latitudes (>55°N) by Bigelow et al. (2003). We assume that the high agreement not only originates from the high quality of the pollen data set used with respect to taxonomical and temporal harmonization, but also relates to the fact that the biomization method employs updated and harmonized schemes assigning pollen taxa to plant functional types to megabiomes. Furthermore, our reconstruction was performed at the megabiome level, a coarser classification than typical biomes, which somewhat reduces the mismatch between geographically adjacent biomes. For example, the biomes of temperate deciduous forest and cool mixed forest are intermingled in Binney et al. (2017), whereas this mismatch does not exist at the megabiome level of temperate forest. Although some regional-scale biomization studies achieve higher agreement with modern biome distributions, such as 97.5% in the Congo Basin by Lebamba et al. (2009), these often rely on more localized datasets with tailored taxa-PFT-biome schemes. As a result, we argue that the data quality as well as the higher spatial and temporal coverage compared to previous biomization studies (Bigelow et al., 2003; Marinova et al., 2018) make our pollen-based megabiome reconstruction a robust dataset for various applications, such as global-scale evaluation of paleo-simulations from Earth System Models (ESMs)."

Reviewer comment: (9) *In conclusion, the paper has potentialities, but it needs more work to be a good contribution to the discipline.*

Response: We sincerely appreciate the reviewer's thoughtful feedback and recognition of the potential of our work. We have carefully addressed the issues raised in the review, please refer to the previous responses.

2. Specific comments

Reviewer comment: (1) *Section 2.3: how are interpolated the biomes in sites from the grid of the ESM (the dots of Fig 1B)?*

Response: Thank you for your comment. We have revised the text in Section 2.3:

Revised text: (in red)

"We assigned the simulated megabiome data taken from the grid-cells where the records are located to each record, and we only considered records and timeslices for which reconstructions are available."

Reviewer comment: (2) *230-233: I disagree that there are no systematic mismatch: there are in the Med region, in the subarctic one ... Note also that TUND and STEP have low agreement (50%)*

Response: Thank you for your comment. The sentence has been removed, and the lower agreement of TUND and STEP has been clarified in the revised text.

Revised text:

"The low taxonomic resolution could also cause mismatches between neighboring forest megabiomes, as well as between tundra (TUND) and grassland (STEP). Woody taxa have been harmonized to the genus level rather than the species level, while herbaceous taxa are generally harmonized to the family level, except for common taxa like *Artemisia*, *Thalictrum*, and *Rumex.* This reduces the ecological information available for PFT assignment (Chen et al., 2010).

[...]

Similarly, TUND may have been misrepresented as STEP on the Tibetan Plateau. This misrepresentation can be attributed to they share dominant characteristic species of Poaceae and Cyperaceae, whereas STEP is defined by fewer PFTs and therefore preferentially allocated. However, the woody PFTs are generally not defined in STEP, leading to a potential misallocation to TUND rather than STEP in cases of woody pollen grain occurrences (from long-distance transportation or local existence) in open landscape samples (Marinova et al., 2018; Chen et al., 2010), such as mismatches in southern Europe."

Reviewer comment: (3) *233-235: I think that the good agreement comes to the fact that the comparison is restricted to 8 megabiomes. In the previous papers, finer biomes are considered and often the mismatch is between climatically neighbor biomes (this is less possible with megabiomes).*

Response: We agree that the good agreement is partly due to being limited to 8 megabiomes. We have revised the text accordingly.

Revised text:

"We consider global-scale, pollen-based megabiome reconstructions reliable because record-by-record comparisons of reconstructed megabiomes at timeslice 0 cal. ka BP from 2,232 available records with modern potential megabiomes indicate an 80.2% agreement (Table 3). This consistency exceeds that reported in previous large-scale biomization studies validated against modern biome distributions, such as the 53% agreement in Arctic high-latitudes (>55°N) by Bigelow et al. (2003). We assume that the high agreement not only originates from the high quality of the pollen data set used with respect to taxonomical and temporal harmonization, but also relates to the fact that the biomization method employs updated and harmonized schemes assigning pollen taxa to plant functional types to megabiomes. Furthermore, our reconstruction was performed at the megabiome level, a coarser classification than typical biomes, which somewhat reduces the mismatch between geographically adjacent biomes. For example, the biomes of temperate deciduous forest and cool mixed forest are intermingled in Binney et al. (2017), whereas this mismatch does not exist at the megabiome level of temperate forest. Although some regional-scale biomization studies achieve higher agreement with modern biome distributions, such as 97.5% in the Congo Basin by Lebamba et al. (2009), these often rely on more localized datasets with tailored taxa-PFT-biome schemes."

Reviewer comment: (4) *259-263: I do not understand this sentence.*

Response: We apologize for the lack of clarity in the original text. The sentence has been revised for better understanding as follows:

Revised text:

"Similarly, TUND may have been misrepresented as STEP on the Tibetan Plateau. This misrepresentation can be attributed to they share dominant characteristic species of Poaceae and Cyperaceae, whereas STEP is defined by fewer PFTs and therefore preferentially allocated. However, the woody PFTs are generally not defined in STEP, leading to a potential misallocation to TUND rather than STEP in cases of woody pollen grain occurrences (from long-distance transportation or local existence) in open landscape samples (Marinova et al., 2018; Chen et al., 2010), such as mismatches in southern Europe."

Reviewer comment: (5) *Section 3.1.2: It is necessary to try to explain the discrepancies between simulated biomes and potential vegetation by over or under-simulation of some climate variables.*

Response: We have incorporated an additional evaluation using modern climate data as follows:

Revised text:

"The agreement between modern potential megabiomes and simulated megabiomes at timeslice 0 cal. ka BP is higher for the ESM-representative megabiome (cf. Sect. 2.3) than for individual ESM-based simulation (64.1% vs. 20.0–60.2%; Table 3). As a result, the ESM-representative megabiome depicts more reliable patterns of megabiome dynamics and distribution than individual simulations, with higher agreement especially in Alaska, the Iberian Peninsula, the Alps, the Atlantic Coastal Plain of North America, and the southeastern United States (Fig. 1 and Fig. A2). However, there are still certain regions with low agreement, probably due to climatic biases. These include nearly all highlands (such as the central-southern Rockies, the central Andes, and the Tibetan Plateau) for which an overestimation of the temperature can be expected in the models due to a much lower orography than in reality caused by the smoothing in the coarse spatial resolution (3.75°x3.75° and 5°x5°) of the model grids (Fig. A3a–b). All models simulate non-forest megabiomes instead of forest in the Mediterranean region, which can be attributed to the models simulating a climate that is too seasonally dry, with, for example, too-warm summers and too-dry winters (Fig. A3a, d). The TRACE-21K simulation as well as the MPI-ESM simulations fail to reproduce the boreal forest (BOFO) in Alaska, which is then also reflected in the ESM-representative megabiomes. This failure is likely due to the simulated climate being too cold in this region, preventing the establishment of boreal forests under modeled conditions (Fig. A3a, d). Similar to the reconstructions, the transition zones between temperate forest (TEFO) and non-forest megabiomes, such as the East Asian summer monsoon margin, are regions with lower simulated megabiome agreement to the modern potential megabiome distribution. In North Africa, the models also tend to underestimate the northern extension of the grassland and dry shrubland (STEP) and incorrectly assign (warm) savanna and dry woodland (SAVA) records to tropical forest (TRFO). This is related to the biomization procedure for the model results that only relies on simulated vegetation cover fractions and simulated climate, whereas savannas are additionally determined by other ecological processes such as fire intensity and frequency (Dallmeyer et al., 2019) or grazing (van Langevelde et al., 2019)."

Reviewer comment: (6) *Figure 3: how is obtained the ice-sheet extension data from pollen?*

Response: Thank you for your comment. To clarify, the ice-sheet extension data in Figure 3 are not derived from pollen data. Instead, they are based on an ice-sheet ensemble set designed for fair comparisons among simulations.

Text: "We also created an ice-sheet ensemble set with a spatial resolution of 3.75°, synthesized from the maximum extent of ICE-5G, ICE-6G, and GLAC-1D reconstructions, for fair comparisons among simulations."

Reviewer comment: (7) *320-331: Elenga et al 2000 and Prentice et al 2000 reconstructed steppes at 21 ka BP, which seems more realistic than tundra. Prentice et al 2000 is not cited despite the fact that they produced a full global reconstruction for the 21 ka BP period (and also mid-Holocene). It is a major paper that the authors cannot ignore. 325-332: the problem of assigning TUND to STEP and vice versa should also be discussed; previous papers reconstruct mainly STEP to the Med region and not TUND as here. There are also biases in the reconstructed biomes.*

Response: Thanks for your recommendation and we have included this reference. Our reconstruction may have misrepresented STEP as TUND, which has been discussed in the revised text:

Revised text:

"The low taxonomic resolution could also cause mismatches between neighboring forest megabiomes, as well as between tundra (TUND) and grassland (STEP). Woody taxa have been harmonized to the genus level rather than the species level, while herbaceous taxa are generally harmonized to the family level, except for common taxa like *Artemisia*, *Thalictrum*, and *Rumex*. This reduces the ecological information available for PFT assignment (Chen et al., 2010).

[...]

Similarly, TUND may have been misrepresented as STEP on the Tibetan Plateau. This misrepresentation can be attributed to they share dominant characteristic species of Poaceae and Cyperaceae, whereas STEP is defined by fewer PFTs and therefore preferentially allocated. However, the woody PFTs are generally not defined in STEP, leading to a potential misallocation to TUND rather than STEP in cases of woody pollen grain occurrences (from long-distance transportation or local existence) in open landscape samples (Marinova et al., 2018; Chen et al., 2010), such as mismatches in southern Europe."

Reviewer comment: (8) *333-334: TUND seems more extended in Europe at 16-13ka than at 21ka, while the warming has already started, why?*

Response: Thank you for your comment. We have clarified this in the revised text:

Revised text:

"Deglaciation (represented by the timeslices 16 and 13 cal. ka BP): Compared with the LGM, the extratropical megabiomes experienced a remarkable expansion to higher latitudes that coincided with the retreat of the continental ice sheets (Fig. 3). In particular, BOFO, TUND, and TEFO underwent a more extensive expansion compared to the other megabiomes in both our reconstructions and simulations; a result similar to previous biomization studies (such as Binney et al., 2017 and Cao et al., 2019 in north of 30°N). "

Reviewer comment: (9) *342-353: Biomization starts to be realistic in the Holocene, much more than for the cold periods. But It is strange that there is no Mediterranean vegetation (WTFO) in the Med area during all the Holocene, as well for simulation as reconstruction. For simulation WTFO is replaced by STEP and for reconstruction, it is by TEFO.*

Response: (a) All models consistently simulate steppe (STEP) rather than subtropical forest (WTFO) in the Mediterranean region during all the Holocene, likely due to systematic biases in the simulated climate, as described in the revised text.

Revised text: "The Mediterranean region has warm to hot dry summers and mild wet winters. Modeling studies report systematic model biases of too-warm summers and too-dry winters in this region (García-Herrera and Barriopedro, 2018). A comparison with modern data shows similar climate biases in the simulations which may indicate similar systematic biases in the past. This would explain the underrepresentation of the cover fraction of woody PFTs in the simulations (Fig. A3a, d).**"**

(b) In our reconstruction, we may have misrepresented WTFO as TEFO, which has been clarified in the revised text.

Revised text:

"The low taxonomic resolution could also cause mismatches between neighboring forest megabiomes, as well as between tundra (TUND) and grassland (STEP). Woody taxa have been harmonized to the genus level rather than the species level, while herbaceous taxa are generally harmonized to the family level, except for common taxa like *Artemisia*, *Thalictrum*, and *Rumex*. This reduces the ecological information available for PFT assignment (Chen et al., 2010). For instance, different species within *Pinus*, *Alnus*, *Fagus*, and *Betula* (Tian et al., 2018) have different bioclimatic controls, phenology, and life forms, but identification at the genus level results in them being shared by key PFTs in different forest megabiomes (e.g., WTFO vs. TEFO, TEFO vs. BOFO) when assigning taxa to PFTs. One of the typical areas in which this problem occurs is southern Scandinavia. Pollen grains from *Betula pendula* in temperate forests and *Betula pubescens* in boreal forests (Beck et al., 2016) in this region can only be identified to genus level, resulting in these two key species not being able to serve as indicators to distinguish between temperate and boreal forests.**"**

(c) However, the WFTO is not the dominant potential natural megabiome in the modern Mediterranean region. As shown in the figure below, which depicts the distribution of modern potential natural megabiomes aggregated from modern potential natural vegetation (Ramankutty and Foley, 1999; Ramankutty et al., 2010). It represents the world's vegetation cover that had most likely existed for 1986–1995 C.E. in equilibrium with present-day climate and natural disturbance in the absence of human activities.
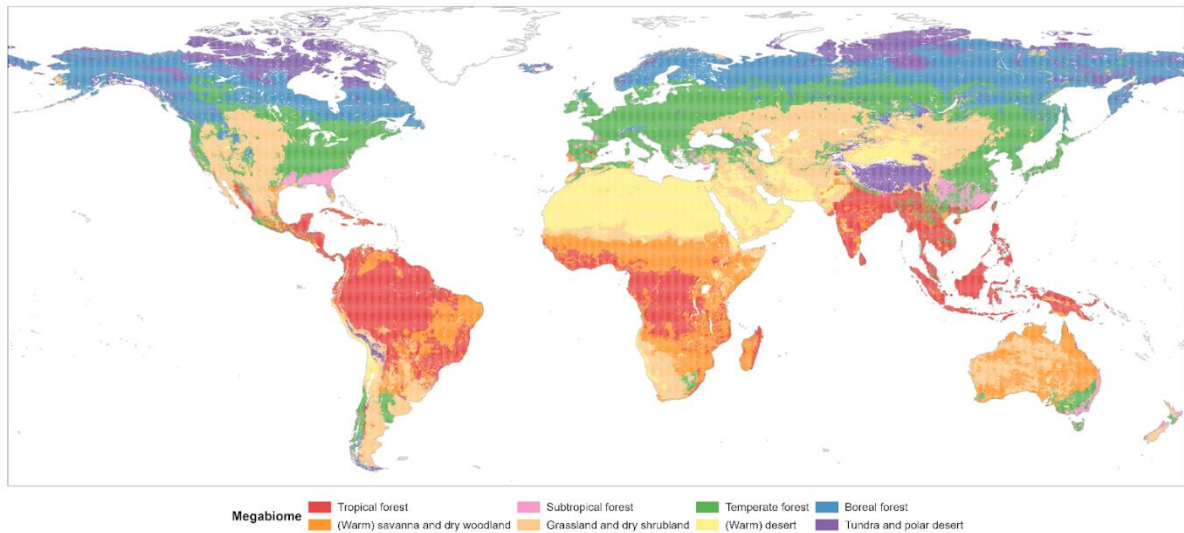
Figure. Spatial patterns of modern potential natural megabiome distributions.

Reviewer comment: (10) *356-361: I read on the maps the opposite to what is claimed: I see a forest degradation in the simulations (maps at right) not in the reconstructions (maps at left). In the whole Holocene, reconstructions show a constant TEFO, while simulation shows steppes in the second part of the Holocene, tending to show that there is a bias towards aridity in the models. Pollen reconstructions may sometimes be influenced by human deforestation, but it has been shown in previous papers that biomization is robust as regards as this disturbance.*

Response: We have revised this paragraph based on your suggestion.
Revised text:

"Mid-Holocene to Late Holocene (represented by the timeslices 6 and 3 cal. ka BP): The spatial patterns of megabiome distributions during this period are only slightly different from those of the early Holocene. TRFO, for example, expanded in Mesoamerican reconstructions and simulations. It is also worth noting that the forest megabiomes have not obviously shifted since the Late Holocene, as revealed by both reconstructions and simulations. Given that the simulated vegetation was in a quasi-equilibrium with the climate and unaffected by humans, this implies a relatively stable climate in that period. Therefore, we propose that enhanced anthropogenic disturbances over this time period did not promote forest degradation at board spatial scale, and that biomization is robust as regards as this disturbance (Prentice et al., 1996; Gotanda et al., 2008)."

Reviewer comment: (11) *Figure 4: It does not exist the possibility to computed significance levels for EDM?*

Response: Thank you for the comment. While it is technically feasible to compute significance levels for the EMD, we have chosen not to perform such calculations as they are not directly relevant to the goals of our study.

The Earth Mover's Distance (EMD) is designed to quantify the degree of mismatch between reconstructions and simulations by integrating uncertainties and weighted ecological and climatic distances. Its primary purpose is to provide a nuanced, distribution-based metric for evaluating spatiotemporal patterns of biome differences rather than testing for random variation or statistical significance.

Significance levels typically require assumptions about the underlying data distribution or the application of resampling techniques, such as bootstrapping. However, the weighting scheme we employed in our EMD calculations is highly context-specific, reflecting ecological and climatic gradients rather than stochastic variation. This specificity makes conventional significance testing less informative in our context.

Moreover, the core objective of our analysis is to assess the relative agreement between reconstructions and simulations across regions and timeslices, focusing on pattern visualization and identifying areas with stronger or weaker model-data agreement. Introducing significance levels would not substantially contribute to the interpretability or scientific value of our results, as the threshold would be inconsistent across different spatial and temporal contexts due to variations in data density and spatial autocorrelation.

To ensure methodological clarity, we emphasize the robustness of our approach by employing carefully defined affinity scores and weighting schemes that directly address vegetation-climate dynamics. These components have been selected to ensure that the EMD reflects biologically and climatically meaningful differences, aligning with the central aims of our study.

Reviewer comment: (12) *Caption of Fig.4: The sentence "The largest datamodel deviations occur during the LGM and early deglaciation periods" should not be put in the caption.*

Response: We've removed it as you suggested.

Reviewer comment: (13) *385-389: "the best data-model agreement occurs during the Bølling-Allerød interstadial (represented by the timeslice 14 cal. ka BP)": This appears to be true with the global EMD but in the regional EMD, it does not appear that 14ka EDM was minimum in any regions.*

Response: The global data-model EMD at each timeslice is derived from the median EMDs of clustered regions at that timeslice. While the global EMD indicates the best data-model agreement during the Bølling-Allerød interstadial (14 cal. ka BP), this does not necessarily mean that the EMD for this timeslice was the minimum in any specific region. Instead, the global pattern represents the aggregated dynamics across all regions, where regional variations may exhibit significant heterogeneity. This distinction reflects the synthesis approach used in our analysis and highlights the importance of considering both global and regional perspectives.

Reviewer comment: (14) *Section 3.3: it should be useful to summarize by a table the biases found in simulations and giving an interpretation of which climate variable is responsible of the biases*

Response: We do not have pollen-independent climate reconstructions. So we cannot really say when and why the models have a climate that is too different, causing differences in vegetation compared to the reconstructions. There are some other reconstructions available, but it is not always clear which proxy records which climate variable, and deciding this is beyond our expertise, and a fair comparison of the past model climate with the reconstructions would be a huge effort and beyond the scope of this paper. Therefore, we can only compare the modern observed climate with the models and add a new table to the appendix.

New table:

Table A1. The median difference in bioclimatic variables between ESM-based simulations at 0 cal. ka BP and observations by regions. The regional clustering is shown in Figure 4b. Bioclimatic variables: T_{warm} - mean temperature of warmest quarter, T_{cold} - mean temperature of coldest quarter, P_{warm} - precipitation of warmest quarter, and P_{cold} - precipitation of coldest quarter. A positive sign in the simulation ensemble difference indicates that the number of simulations that overestimate the bioclimatic variable is greater than the number that underestimate it among the six simulations, while a negative sign indicates the opposite, and positive/negative signs indicate that they are equivalent. Confidence among the six simulations is indicated by one, two, and three asterisks for four, five, and six simulations sharing the same sign, respectively.

Regions	Bioclimatic variables	MPI-ESM		CLIMBER-X		CCSM3		Simulation ensemble	
		MPI-ESM_GLAC1D	MPI-ESM_ICE6G	CLIMBER-X_GLAC1D	CLIMBER-X_ICE6G	TRACE-21K-I_ICE5G	TRACE-21K-II_ICE5G	Difference	Confidence
Asia 1	T_{warm}	-2.4	-2.3	3.1	3.6	-10.0	-5.5	-	*
	T_{cold}	1.9	3.1	4.3	4.7	-10.4	-9.9	+	*
	P_{warm}	36.3	26.6	43.9	52.2	-25.9	-1.1	+	*
	P_{cold}	10.5	12.1	26.6	23.7	-3.0	1.8	+	**
North America 1	T_{warm}	-0.4	-0.4	0.8	0.7	-8.2	-5.4	-	*
	T_{cold}	3.3	4.7	3.4	3.7	-6.3	-5.5	+	*
	P_{warm}	81.9	55.4	24.7	29.7	-12.7	12.4	+	**
	P_{cold}	31.6	30.7	64.8	58.8	18.7	20.7	+	***
Asia 2	T_{warm}	-3.8	-3.6	1.7	2.0	-8.2	-4.1	-	*
	T_{cold}	2.7	3.3	6.6	6.4	-5.2	-3.2	+	*
	P_{warm}	34.7	19.0	14.2	38.3	-71.3	-53.3	+	*
	P_{cold}	29.3	25.9	-14.2	-6.7	6.7	13.6	+	*
Asia 4	T_{warm}	0.7	1.1	-0.8	0.6	2.0	2.4	+	**
	T_{cold}	0.4	0.5	-0.3	0.4	-2.3	-2.5	+/-	
	P_{warm}	0.7	-25.3	-33.0	-33.5	-4.3	-0.6	-	**
	P_{cold}	51.2	49.5	-0.4	1.1	70.3	79.4	+	**
Africa 1	T_{warm}	-0.6	-0.2	0.8	-0.5	-1.8	-1.0	-	**
	T_{cold}	-0.8	-1.2	1.9	0.8	-1.5	-1.0	-	*
	P_{warm}	-13.6	-13.6	23.7	24.3	2.8	5.4	+	*
	P_{cold}	-0.5	-0.5	13.2	14.0	-2.3	-2.4	-	*
Europe 2	T_{warm}	0.9	0.5	0.0	-0.9	-2.8	-1.0	-	*
	T_{cold}	1.1	1.4	2.0	1.3	-0.5	0.8	+	**
	P_{warm}	-17.9	-18.1	38.8	36.4	-5.6	-4.6	-	*
	P_{cold}	-26.2	-52.3	-44.0	-43.6	-46.4	-42.9	-	***
North America 3	T_{warm}	0.1	0.3	-0.4	0.0	-1.7	-1.0	+/-	
	T_{cold}	-0.7	-0.6	1.3	0.6	-1.5	-0.7	-	*
	P_{warm}	-18.2	-15.5	31.4	21.3	-23.1	-12.2	-	*
	P_{cold}	50.3	52.9	5.2	8.1	39.6	47.4	+	***
South America 2	T_{warm}	-1.8	-1.7	-3.3	-3.1	-0.7	-0.3	-	***
	T_{cold}	-2.1	-1.7	-2.9	-3.0	-1.6	-1.7	-	***
	P_{warm}	125.6	96.7	-45.1	57.0	7.7	14.7	+	**
	P_{cold}	-13.5	2.1	-0.8	-0.9	-75.3	-89.7	-	**
Indo-Pacific 1	T_{warm}	-1.0	-1.2	0.7	0.3	0.5	0.0	+/-	
	T_{cold}	0.6	0.7	0.4	0.1	0.2	0.0	+	**
	P_{warm}	30.6	-16.8	-155.9	-151.5	16.7	-16.0	-	*
	P_{cold}	-45.2	-38.8	-89.8	-87.2	-119.1	-106.4	-	***

Reviewer comment: (15) *Section 4 (conclusions): this conclusion is short and superficial. What is the origin of the ESM biases according to reconstruction at least for the main ones? Has this paper filled the initial objectives?*

Response: Thank you for your comment, we have revised the conclusion.

Revised text:

"This study presents a global megabiome reconstruction for 43 timeslices at 500-year intervals over the past 21,000 years, based on the most extensive taxonomically and temporally standardized fossil pollen dataset. The dataset's reliability is supported by a high agreement (~80%) with modern potential natural vegetation, and its general consistency with the paleosimulation ensemble further underscores its robustness for exploring past biome dynamics. With its high temporal and spatial coverage, it offers an unprecedented resource, not only for exploring long-term vegetation dynamics and their drivers, but also for diverse research contexts, including paleoclimate, biodiversity, and land-use studies. Furthermore, the dataset supports the evaluation of ESM-based paleo-megabiome simulations and offer insights for identifying potential biases in climate models. Its consistent structure and broad applicability allow us to advance our integrative understanding of past, present, and future Earth system dynamics."

References:

Beck, P., Caudullo, G., de Rigo, D., and Tinner, W.: Betula pendula, Betula pubescens and other birches in Europe: distribution, habitat, usage and threats, in: European Atlas of Forest Tree Species, edited by: San-Miguel-Ayanz, J., de Rigo, D., Caudullo, G., Houston Durrant, T., Mauri, A., Publication Office of the European Union, Luxembourg, 70–73, https://doi.org/ 10.7892/boris.80789, 2016.

Bigelow, N. H., Brubaker, L. B., Edwards, M. E., Harrison, S. P., Prentice, I. C., Anderson, P. M., Andreev, A. A., Bartlein, P. J., Christensen, T. R., Cramer, W., Kaplan, J. O., Lozhkin, A. V., Matveyeva, N. V., Murray, D. F., McGuire, A. D., Razzhivin, V. Y., Ritchie, J. C., Smith, B., Walker, D. A., Gajewski, K., Wolf, V., Holmqvist, B. H., Igarashi, Y., Kremenetskii, K., Paus, A., Pisaric, M. F. J., and Volkova, V. S.: Climate change and Arctic ecosystems: 1. Vegetation changes north of 55°N between the last glacial maximum, mid-Holocene, and present, J. Geophys. Res. Atmos., 108, https://doi.org/10.1029/2002JD002558, 2003.

Binney, H., Edwards, M., Macias-Fauria, M., Lozhkin, A., Anderson, P., Kaplan, J. O., Andreev, A., Bezrukova, E., Blyakharchuk, T., Jankovska, V., Khazina, I., Krivonogov, S., Kremenetski, K., Nield, J., Novenko, E., Ryabogina, N., Solovieva, N., Willis, K., and Zernitskaya, V.: Vegetation of Eurasia from the last glacial maximum to present: Key biogeographic patterns, Quat. Sci. Rev., 157, 80–97, https://doi.org/10.1016/j.quascirev.2016.11.022, 2017.

Cao, X., Tian, F., Dallmeyer, A., and Herzschuh, U.: Northern Hemisphere biome changes (>30°N) since 40 cal ka BP and their driving factors inferred from model-data comparisons, Quat. Sci. Rev., 220, 291–309, https://doi.org/10.1016/j.quascirev.2019.07.034, 2019.

Chen, Y., Ni, J., and Herzschuh, U.: Quantifying modern biomes based on surface pollen data in China, Global Planet. Change, 74, 114–131, https://doi.org/10.1016/j.gloplacha.2010.09.002, 2010.

Cheng, W., Dan, L., Deng, X., et al.: Global monthly gridded atmospheric carbon dioxide concentrations under the historical and future scenarios, Sci. Data, 9, 83, https://doi.org/10.1038/s41597-022-01196-7, 2022.

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

Davis, B. A. S., Fasel, M., Kaplan, J. O., Russo, E., and Burke, A.: The climate and vegetation of Europe, northern Africa, and the Middle East during the Last Glacial Maximum (21 000 yr BP) based on pollen data, Clim. Past, 20, 1939–1988, https://doi.org/10.5194/cp-20-1939-2024, 2024.

Duan, R., Huang, G., Wang, F., Tian, C., and Wu, X.: Observations over a century underscore an increasing likelihood of compound dry-hot events in China, Earth's Future, 12, e2024EF004546, https://doi.org/10.1029/2024EF004546, 2024.

Elenga, H., Peyron, O., Bonnefille, R., Jolly, D., Cheddadi, R., Guiot, J., Andrieu, V., Bottema, S., Buchet, G., De Beaulieu, J, L., Hamilton, A. C., Maley, J., Marchant, R., Perez-Obiol, R., Reille, M., Riollet, G., Scott, L., Straka, H., Taylor, D., Van Campo, E., Vincens, A., Laarif, F., and Jonson, H.: Pollen-based biome reconstruction for southern Europe and Africa 18,000 yr bp, J. Biogeogr., 27, 621–634, https://doi.org/10.1046/j.1365-2699.2000.00430.x, 2000.

Gotanda, K., Nakagawa, T., Tarasov, P. E., and Yasuda, Y.: Disturbed vegetation reconstruction using the biomization method from Japanese pollen data: Modern and Late Quaternary samples, Quat. Int., 184, 56–74, https://doi.org/10.1016/j.quaint.2007.09.030, 2008.

Harris, I., Osborn, T.J., Jones, P. and Lister, D.: Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset, Sci. Data 7, 109. https://doi.org/10.1038/s41597-020-0453-3, 2020.

Lebamba, J., Ngomanda, A., Vincens, A., Jolly, D., Favier, C., Elenga, H., and Bentaleb, I.: Central African biomes and forest succession stages derived from modern pollen data and plant functional types, Clim. Past, 5, 403–429, https://doi.org/10.5194/cp-5-403-2009, 2009.

Marinova, E., Harrison, S. P., Bragg, F., Connor, S., de Laet, V., Leroy, S. A. G., Mudie, P., Atanassova, J., Bozilova, E., Caner, H., Cordova, C., Djamali, M., Filipova-Marinova, M., Gerasimenko, N., Jahns, S., Kouli, K., Kotthoff, U., Kvavadze, E., Lazarova, M., Novenko, E., Ramezani, E., Röpke, A., Shumilovskikh, L., Tanţău, I., and Tonkov, S.: Pollen-derived biomes in the Eastern Mediterranean–Black Sea–Caspian-Corridor, J. Biogeogr., 45, 484–499, https://doi.org/10.1111/jbi.13128, 2018.

Nolan, C., Overpeck, J. T., Allen, J. R. M., Anderson, P. M., Betancourt, J. L., Binney, H. A., Brewer, S., Bush, M. B., Chase, B. M., Cheddadi, R., Djamali, M., Dodson, J., Edwards, M. E., Gosling, W. D., Haberle, S., Hotchkiss, S. C., Huntley, B., Ivory, S. J., Kershaw, A. P., Kim, S, H., Latorre, C., Leydet, M., Lézine, A, M., Liu, K, B., Liu, Y., Lozhkin, A. V., McGlone, M. S., Marchant, R. A., Momohara, A., Moreno, P. I., Müller, S., Otto-Bliesner, B. L., Shen, C., Stevenson, J., Takahara, H., Tarasov, P. E., Tipton, J., Vincens, A., Weng, C., Xu, Q., Zheng, Z., and Jackson, S. T.: Past and future global transformation of terrestrial ecosystems under climate change, Science, 361, 920–923, https://doi.org/10.1126/science.aan5360, 2018.

O'Donnell, M. S. and Ignizio, D. A.: Bioclimatic predictors for supporting ecological applications in the conterminous United States, U.S. Geological Survey Data Series, 691, 1–10, 2012.

Prentice, C., Guiot, J., Huntley, B., Jolly, D., and Cheddadi, R.: Reconstructing biomes from palaeoecological data: a general method and its application to European pollen data at 0 and 6 ka, Clim. Dyn., 12, 185–194, https://doi.org/10.1007/BF00211617, 1996.

Prentice, I.C., Jolly, D. and Biome 6000 Participants: Mid-Holocene and glacial-maximum vegetation geography of the northern continents and Africa, J. Biogeogr., 27, 507–519, https://doi.org/10.1046/j.1365-2699.2000.00425.x, 2000.

Ramankutty, N. and Foley, J. A.: Estimating historical changes in global land cover: Croplands from 1700 to 1992, Global Biogeochem. Cycles, 13, 997–1027, https://doi.org/10.1029/1999GB900046, 1999.

Ramankutty, N., Foley, J. A., Hall, F. G., Collatz, G. J., Meeson, B. W., Los, S. O., Brown De Colstoun, E., and Landis, D. R.: ISLSCP II Potential Natural Vegetation Cover, ORNL DAAC [data set], https://doi.org/10.3334/ORNLDAAC/961, 2010.

Tian, F., Cao, X., Dallmeyer, A., Lohmann, G., Zhang, X., Ni, J., Andreev, A., Anderson, P. M., Lozhkin, A. V., Bezrukova, E., Rudaya, N., Xu, Q., and Herzschuh, U.: Biome changes and their inferred climatic drivers in northern and eastern continental Asia at selected times since 40 cal ka bp, Veg. Hist. Archaeobot., 27, 365–379, https://doi.org/10.1007/s00334-017-0653-8, 2018.

van Langevelde, F., van de Vijver, C. A. D. M., Prins, H. H. T., and Groen, T. A.: Effects of grazing and browsing on tropical savanna vegetation, in: The Ecology of Browsing and Grazing II, edited by: Gordon, I. J. and Prins, H. H. T., Springer International Publishing, Cham, Switzerland, 237–257, https://doi.org/10.1007/978-3-030-25865-8_10, 2019.