Harmonizing plant functional type distributions for evaluating Earth System Models

Anne Dallmeyer[1], Martin Claussen[1,2] and Victor Brovkin[1]

[1]Max Planck Institute for Meteorology, Bundesstrasse 53, 20146 Hamburg, Germany
5 [2]Meteorological Institute, Centrum für Erdsystemforschung und Nachhaltigkeit (CEN), Universität Hamburg, Bundesstrasse 55, 20146 Hamburg, Germany

Correspondence to: Anne Dallmeyer (anne.dallmeyer@mpimet.mpg.de)

Abstract.

Dynamic vegetation models simulate global vegetation in terms of fractional coverages of a few plant functional types
10 (PFTs). Although these models often share the same concept, they differ with respect to the number and kind of PFTs, complicating the comparability of simulated vegetation distributions. Pollen-based vegetation reconstructions are initially only available in form of time-series of individual taxa that are not distinguished in the models. Thus, to evaluate simulated vegetation distributions, the modelling results and pollen-based vegetation reconstructions have to be converted into a comparable format. The classical approach is the method of biomisation, but hitherto, PFT-based biomisation methods were
15 only available for individual models. We introduce and evaluate a simple, universally applicable technique to harmonize PFT-distributions by assigning them into nine mega-biomes, using only assumptions on the minimum PFT cover fractions and few bioclimatic constraints (based on the 2m-temperature). These constraints mainly follow the limitation rules used in the classical biome models (here: BIOME4). We test the method for six state-of-the-art dynamic vegetation models that are included in Earth System Models based on pre-industrial, mid-Holocene and Last Glacial Maximum simulations. The
20 method works well, independent of the spatial resolution or the complexity of the models. Large biome belts (such as tropical forest) are generally better represented than regionally confined biomes (warm-temperate forest, savanna). The comparison with biome distributions inferred via the classical biomisation approach of forcing biome models (here: BIOME1) with the simulated climate states shows that the PFT-based biomisation is even able to keep up with the classical method. However, as the new method considers the PFT distributions actually calculated by the Earth System Models, it
25 allows for a direct comparison and evaluation of simulated vegetation distributions which the classical method cannot do. Thereby, the new method provides a powerful tool for the evaluation of Earth System Models in general.

1 Introduction

Within dynamic global vegetation models (DGVM), the natural vegetation distribution is usually represented in the form of plant functional types (PFT), i.e. plants are grouped with regard to their physiology and physiognomy (Prentice et al. 2007).
30 These PFTs differ with respect to phenology, albedo, morphological and photosynthetic parameters and are usually constraint

by an individual bioclimatic range of tolerance defined by temperature thresholds. These thresholds represent the cold resistance, chilling and heat requirements of the plants and determine the area where the PFTs can establish.

In most DGVMs, a 'mosaic' approach is used, i.e. each grid-box of the land surface is split into separate parts for a non-vegetated and a vegetated fraction that is further tiled in mosaics, taking sub-grid scale heterogeneity into account. Thus, several PFTs can cover the same grid-cell and compete for space via their net primary productivity (e.g. Sitch et al., 2003; Krinner et al., 2005; Reick et al., 2013). Non-vegetated area (seasonally bare soil or permanently bare ground) is produced where plant productivity is too low.

Although the main principles for the calculation of PFT-distributions are similar among most DGVMs, they vary regarding the number and kind of PFTs used to represent the global vegetation. Natural (non-anthropogenic) PFTs range from two in e.g. VECODE to ten in e.g. LPJ and ORCHIDEE (Tab.1). Even within the same model, PFT variety can differ between individual simulations, e.g. due to the inclusion of land-use types. These differences among the simulations and models prohibits the inter-model comparability of simulated global vegetation distributions and the comparability with pollen-based biome reconstructions.

Pollen records are originally displayed in form of pollen percentages or pollen accumulation rates, that can not be directly compared to plant functional type distributions, as pollen records do not reflect the actual plant abundances. For a systematic comparison of simulated plant functional type distributions and reconstructions, both need to be converted in a compatible format. In the last two decades, taxa to PFT assignment-methods and the method of 'biomisation' for pollen-based reconstructions have been developed (e.g. Prentice et al. 1996, Ni et al., 2010, Harrison et al. 2010), so that pollen assemblages can be grouped into biomes (e.g. tropical forest, temperate steppe, desert). Pollen-based biome syntheses have been provided (Prentice et al., 1998 and 2000, Bigelow et al., 2003; Ni et al. 2010, Harrison 2017, Tian et al. 2017) that have extensively been used to evaluate simulated biome distributions obtained from diagnostic biome models such as BIOME1 or BIOME4 (e.g. Prentice et al., 1992, Haxeltine and Prentice, 1996, Kaplan et al., 2003). These biome models can be forced by observed or simulated climate fields and calculate biome distributions in equilibrium to this input climate. Using this classical method of biomisation, fundamental palaeo-vegetation analysis can be undertaken (e.g. Jolly et al, 1998; Harrison et al., 2003; Wohlfahrt et al., 2008; Harrison et al., 2016, Dallmeyer et al., 2017) without requiring an explicit calculated vegetation distribution by the Earth System Models (ESMs). On the other hand this also means, that existing simulated plant functional type distributions calculated by the DGVMs being dynamically coupled in these models are neglected, since only the simulated climate pattern is taken into account. The biomisation via diagnostic biome models did not include any information on the original PFT-distribution simulated by the Earth System Models. As the DGVMs are generally more complex than the biome models and include more relevant processes, valuable information included in the PFT-distribution gets lost in the classical biomisiation by the biome models. A more appropriate way of biomisation would be to use directly the PFT-distributions calculated by the DGVMs.

Several model studies have taken up this problem by introducing methods for biomising PFT-distributions simulated by DGVMs. Schurgers et al. (2006) derive biomes maps for the Eemian and mid-Holocene from the relative fractional coverage

of the individual PFTs and the soil temperature, both simulated by LPJ. With this method, reconstructed major biome shifts could be reproduced. Roche et al. (2007) used the dominant PFT and the bioclimate limits defined in the biome model BIOME1 (Prentice et al., 1992) to biomise PFT cover fractions for the Last Glacial Maximum (LGM) simulated by VECODE. As VECODE distinguish as main PFTs only trees and herbaceous plants, not all biome types defined in BIOME1 could be considered (e.g. no shrubs). The computed biome map shows reasonable agreement with LGM land cover reconstructions. A similar approach was chosen by Handiani et al. (2012 and 2013) for calculating biome distributions at Heinrich Event 1, based on PFT simulations of TRIFFID and the CLM-DGVM. As these models strongly deviate in their PFT classification, they applied different methods for biomisation. For TRIFFID, they first calculated the dominant PFT in each grid cell following the method by Cruxifix et al. (2005) and afterwards used temperature limitation defined in BIOME4 (Kaplan et al., 2003) to assign the dominant PFTs to mega-biomes. For CLM-DGVM, potential dominant PFTs were estimated by adopting the scheme of Schurgers et al. (2006) and biomes were differentiated with the help of temperature limitations that follow the environmental constraints defined in CCSM3 (i.e. the fully coupled model used in their study, including the land and vegetation model CLM-DGVM).

Recently, Prentice et al. (2011) introduced another approach of biomising plant functional type distributions simulated by dynamic vegetation models. In their method, simulated foliage projective cover (FPC) is used to distinguish between desert, grassland/dry shrubland and forest biomes, which are further divided into forest and savanna like biomes through the vegetation height. The assignment to e.g. boreal, temperate or tropical forest/savanna (or parkland) is controlled by the tree-PFT composition. Climate limits are only used to distinguish the tundra biome. This method has successfully been used in several palaeo-vegetation studies (Kageyama et al., 2013; Calvo and Prentice, 2015) using different versions of the LPJ and ORCHIDEE models.

All of these methods have in common that they have been designed for individual models and hence need specific output not necessarily provided by all models. Therefore these methods can not directly be adopted for all existing dynamic vegetation models. A consistent inter-model comparison of the simulated vegetation distribution and an evaluation of the models against reconstructions on biome level is so far not possible.

To harmonize (palaeo)-vegetation distributions simulated by dynamic vegetation models and thereby facilitate the evaluation of Earth System Models (ESMs) and the comparison of model results and biome reconstructions, we developed a biomisation technique that is based on the PFT-distributions simulated by the DGVMs and few input variables and simple differentiation rules. These include bioclimatic constraints using near-surface air temperature and assumptions on maximum required PFT coverage. We test this method on pre-industrial, mid-Holocene and Last Glacial Maximum vegetation simulations performed in nearly all state-of-the-art dynamic vegetation models. The skill of this biomisation approach is quantified via (standard) metrics, by comparing the converted biome maps with estimates of modern potential biome distributions (Ramankutty and Foley, 1999) and pollen-based biome reconstructions (Biome6000 database, Harrison, 2017).

2 Methods

2.1 Biomisation

The plant functional type (PFT) cover fractions simulated by the individual dynamic vegetation models are converted into nine different mega-biomes (Fig.1), using few bioclimatic limits and assumptions on the maximum required coverage of certain PFTs. The aggregation into the mega-biomes is in line with the definitions of the BIOME6000 project (cf. Harrison, 2017) that are also commonly used for grouping pollen-based biome reconstructions. Bioclimatic limits and the differentiation rules basically follow the biome assignment of the BIOME4 model (Kaplan et al. 2003). As input data, only climatological mean growing degree days (GDD0 and GDD5), monthly mean 2m-air temperature, and multi-year mean PFT cover fractions (e.g. averaged over 100 years) are required. The limitation to few climatic rules and few variables needed enables the application of the method to all state-of-the art dynamic vegetation models.

In detail, the PFTs calculated by the respective dynamic vegetation model are aggregated into the groups 'trees', 'woody PFTs' (i.e. shrubs and all tree PFTs), 'grass' and 'desert' which is calculated by one minus the total vegetation. If the model include land-use types, the affected areas are redistributed to the other PFTs by simply scaling up the other PFT fractions proportionally to their ratio of the total natural vegetation. Based on these groups, regions dominated by trees and regions in which the cover fraction of the woody PFTs exceed 25% (with the additional constraint of the total vegetation cover exceeding 50%) are identified, as these are necessary condition for the assignment of the land cover to forest biomes. Afterwards, the groups 'trees' and 'woody PFTs are split into boreal (GDD5 ≤ 900°C), temperate (GDD5 > 900°C) and tropical trees/wood (Tc > 15.5°C) via temperature limits (cf. Tab.2). If any of theses tree-PFTs are simulated directly in the vegetation model (e.g. in LPJ or ORCHIDEE), the original distributions are taken and the PFT-group is assigned to the dominant tree-type, i.e. the tree type that covers the largest fraction of the grid-cell. These PFT-groups (i.e. boreal woody PFTs, temperate trees / woody PFTs, tropical trees, grass and desert) are the first consistent vegetation classification shared by all input simulations, so that model-to-model comparison is also possible on this PFT level.

For the biomisation, the forests are considered first, i.e. regions in which trees or woody PFTs are dominant or cover an area more than 25% are assigned to tropical, temperate and boreal forests according to the PFT-groups. From the temperate forest, warm regions (i.e. GDD5 > 3000°C) revealing a dominant temperate tree fraction are subtracted and assigned to the biome 'warm-temperate forest'. The remaining area is then tested for fulfilling the constraints for the non-forest biomes. First, the savanna and dry woodland region is identified by bioclimatic limitations (GDD5 > 1200°C and Tc > 10°C) and a woody PFT-coverage of at least 25%. The remaining vegetated area is assigned to the biome 'grassland and dry shrublands', if GDD0 exceeds 800°C or to the biome 'tundra', if GDD0 is below 800°C (cf. BIOME4, Kaplan et al. 2003). The non-vegetated area, i.e. regions in which the total vegetation cover is less than 20%, is either assigned to warm or to cold desert, depending on whether the annual mean temperature is above or below 2°C. For the biome 'tundra', only 10% vegetation cover is needed. A flowchart summarizing the details of the PFT-based biomisation is shown for the VECODE model in the Appendix A.

130   We are aware of the simplicity of this approach, calculating the tundra and the grassland and dry shrubland biomes as residual of the non-forested area, not directly depending on the simulated grass PFT-fraction. We decided to attribute main priority on the forested biomes as this is also the strategy commonly used in DGVMs and Biome Models.

To assess the performance of the biomisation based on simulated PFTs, we additionally biomise the simulated climate fields corresponding to the PFT-distributions in each model. This is the conventionally used procedure to biomise GCM or ESM

135   output (further referred to as classical approach or climate-based method). For this purpose, we use the biome model BIOME1 (Prentice et al., 1992) that calculates the biome distribution in equilibrium to the input climate. As forcing, BIOME1 needs the monthly mean climatological precipitation, near-surface temperature and cloudiness, which were taken from each simulation considered in this study, respectively. This classical biomisation approach can only handle climate data as input, the simulated PFT-distributions from the ESMs used in the here introduced PFT-based method are ignored. The

140   original biomes have been grouped into the same mega-biome classification that is used for the PFT-based approach.

2.2 Simulations

Simulations from nearly all state-of-the-art global dynamic vegetation models that are included in Earth System Models have been selected for biomisation. Six different models could be considered (i.e. JSBACH, TRIFFID, ORCHIDEE, SEIB, LPJ and VECODE). Overall, eight simulations for the pre-industrial climate (PI) and vegetation, four for mid-Holocene (6k)

145   conditions and five for Last Glacial Maximum (LGM) conditions have been used (Tab.3). Most of these simulations were performed within CMIP5/PMIP3 under strict simulation and output protocols enabling direct comparison between the models (Braconnot et al., 2011, Taylor et al., 2012). These include the models MPI-ESM-P, IPSL-CM5A-LR, MIROC-ESM and HadGem2-ESM. Further details on the models and simulations are described in the Appendix B.

For the pre-industrial time-slice, two out of eight simulations (MPI-ESM-T63 and IPSL-ESM-T31) were performed with

150   fixed vegetation distribution, but this has essentially no effect on the biomisation procedure. Therefore, we include these simulations in our analysis. Nevertheless, the PFT-based biome-distributions for these simulations are expected to fit better to the references than the other simulations that ran with interactive vegetation.

We emphasize that this study is thought as an introduction and detailed evaluation of a new biomisation method. It is not seen as evaluation of the different vegetation models with respect to the skill of simulating biome or vegetation distributions.

155   For this purpose, the different vegetation models would have to be forced by the same climate state. Such an ensemble including all DGVMs (used here) does not exist. Therefore we had to take individual simulations that all deviate with respect to the prescribed or simulated climate in the coupled models. These differences in the climatic field among the models and between the models and observations (general climate biases) lead - per se – to differences in the simulated vegetation distributions and biases to the reference vegetation distribution.

2.3 Preparing the reference datasets

As reference, we use the estimated global potential natural vegetation map by Ramankutty and Foley (1999, referred to as RF99 in the following), which is a combination of modern satellite-based vegetation observations (i.e. the DISCover land cover dataset) and the vegetation compilation prepared by Haxeltine and Prentice (1996) that has been taken for regions dominated by land use at present-day. The RF99 dataset is available at 5 minute resolution and distinguishes 15 different biome types that are similar to the mega-biome classification used here. Thus, most biomes could directly be assigned to the mega-biome types (Tab C1). The preparation of the RF99 reference dataset is explained in more detail in the Appendix C.

As further reference data, we use pollen-based biome reconstructions that are available for the modern, mid-Holocene and Last Glacial Maximum time-slice within the Biome6000 database (Harrison, 2017). The biome reconstructions have been grouped into the mega-biomes according to the suggestions made by the Biome 6000 project.

Both vegetation datasets are derived for the modern time-slice not exactly corresponding to the pre-industrial period (around 1850 AD) simulated in the models. While the ice-sheet, the topography and the orbital conditions used for the pre-industrial control simulations are prescribed from modern conditions, greenhouse gases are set to pre-industrial values in the models. These differences in e.g. atmospheric CO_2-concentration between the reference datasets and the simulations may lead to small discrepancies in the model. In addition, the references may be disturbed by anthropogenic influences.

As pre-industrial reference climate, climatological monthly mean data of the year 1901-1930 from the University of East Anglia Climate Research Unit Time Series 4.00 (CRU TS4, University of East Anglia, 2017) has been taken. This is the earliest period available. The CRU TS4 reference climate has additionally been used as forcing for the BIOME1 model to provide a 'best guess' for the pre-industrial biome map. We assume that neither the biomisation of simulated climate states (i.e. the classical method) nor the biomisation of simulated PFTs can agree better with any reference than this biome distribution, derived with a highly tuned biome model and the best global climate observation available. Therefore, we use the level of agreement between the CRU TS4 biome map and the RF99 or the Biome6000 reconstructions as target value for our new biomisation method. The reference biome distributions and the CRU TS4-based biome map are displayed in Fig.2.

2.4 Metrics

2.4.1 Kappa Statistic

The Kappa statistic (Cohen, 1960) is a widely used quantitative map-comparison technique that has often been applied for assessing the performance of vegetation simulations (e.g. Monserud and Leemans, 1992; Prentice et al., 1992; Diffenbaugh et al., 2003; Tang et al., 2009). The Kappa statistic not only includes the actual observed similarity (p_0) of two categorical maps, but also considers the expected agreement (p_e), i.e. the agreement by chance. For each pair of compared grid-cells (or a pair of grid-cell and site) taken from the reference and the simulated biome distribution, a confusion matrix is prepared containing all combinations of referenced and simulated biomes. Based on this error matrix, the agreement for each

individual mega-biome is given by the following Eq. (1, taken from Tang et al., 2009):
$$\kappa_i = \frac{p_{ii} - p_{i.r}\, p_{c.i}}{\left((p_{i.r} + p_{c.i})/2 - p_{i.r}\, p_{c.i}\right)}$$

(1)

where p_{ii} is the individual entry for biome i on the main diagonal of the confusion matrix and $p_{i.r}$ and $p_{c.i}$ are the row total and

the column total of each biome i, respectively. The overall agreement is derived by Eq. (2):
$$\kappa = \frac{p_0 - p_e}{1 - p_e}$$
(2)

with $\quad p_0 = \sum_{i=1}^{n} p_{ii} \quad$ and $\quad p_e = \sum_{i=1}^{n} p_{i.r}\, p_{c.i}$. κ ranges from 0 (not better than agreement by chance) to 1 (perfect

agreement). We additionally use the thresholds suggested by Landis and Koch (1977), classifying a κ below 0.4 into poor agreement, values between 0.4 and 0.75 in fair to good agreement and values exceeding 0.75 into very good to excellent agreement.

2.4.2 Fractional Skill Score (FSS)

The standard Kappa statistic underestimates the similarity of maps sharing a similar biome distribution but being slightly offset from each other (Foody, 2002; Tang et al., 2009). This problem is usually overcome by using the Fuzzy Kappa statistic allowing for fuzziness in category and fuzziness in location (Hagen, 2003 and 2009), but the Fuzzy Kappa statistic is only applicable to assess the similarity of categorical maps and can not be used for single point to gridded-data comparison. Biome reconstructions only exist for single sites and usually indicate not only the local nor the regional vegetation, but may contain a large extra-regional component, depending e.g. on the configuration (mainly the size) of the lake (Jacobsen and Bradshaw, 1981). A single grid-cell to point comparison is, thus, only partly meaningful, more advisable is the inclusion of the surrounding grid-cells of the sites. Therefore, we looked for a metric taking agreement in the neighbourhood into account (such as the Fuzzy Kappa statistic) that could easily be adapted to site to gridded-data comparison. We decided to use the fractional skill score (FSS, Roberts and Lean, 2008). While this method was initially developed and applied for expressing the performance of precipitation forecasts (e.g., Gilleland et al., 2009; Mittermaier et al., 2013; Wolff et al., 2014), it has recently been successfully used for different hydrological patterns (Koch et al. 2017). We further adapted the FSS method to biome distributions. For each mega-biome type, the reference (ref) and simulation (sim) is truncated into a binary map, i.e. we construct 18 maps (9 for the reference, 9 for the simulation), in which the grid cell being covered by the respective mega-biomes are filled with the value '1' and all other grid-cells are assigned to the value '0'. Based on these maps, the mean fractional coverage of the respective mega-biome within the neighbourhood N_{ij} (3 grid-cells in each direction for T31, 6 for T63, 1 for 10°grid) of each cell is calculated for the reference and the simulation. Afterwards, the mean-square error (MSE) between the simulation and the reference fractions for each individual mega-biome is calculated and normalized by the MSE

representing the worst-case agreement (MSEw), i.e. the MSE reflecting no similarity between the reference and the

220 simulation. The fractional skill score is then given by Eq.(3) $\quad FSS=1-\dfrac{MSE}{MSEw}$ (3)

with $\quad MSE=\dfrac{1}{N}\sum\limits_{i=1}^{N_i}\sum\limits_{j=1}^{N_j}\left[ref_{ij}-sim_{ij}\right]$ and: $\quad MSEw=\dfrac{1}{N}[\sum\limits_{i=1}^{N_i}\sum\limits_{j=1}^{N_j}ref_{ij}^2+\sum\limits_{i=1}^{N_i}\sum\limits_{j=1}^{N_j}sim_{ij}^2]$; N is the number of all

neighbourhoods.

Following Robert and Lean (2008), we define the lowest skill by the FSS_{ran} of a random biome distribution with the same

fractional coverage as the observed one over the domain (f_0). Likewise, the target skill is given by the FSS that is reached for

225 a uniform distribution of the observed biome fraction everywhere in the domain ($FSS_{uni} = 0.5 + f_0/2$). As FSS_{ran} and FSS_{uni}

deviate between the individual biomes, we compare the relative FSS (rFSS) given by FSS -FSS_{uni}.

The total rFSS is calculated as mean of all individual mega-biome scores. The total and individual rFSS can ranges from ca.

-0.5 (as good as a random distribution) to ca. 0.5 (perfect agreement), depending on the extent of the individual biomes. The

skill to reach is zero for all biomes. For simplicity, we use as abbreviation for the relative FSS also just FSS.

230

2.4.3 Best Neighbour Score

Neither the FSS nor the Fuzzy Kappa statistic is in its original format applicable for the comparison of site-data vs. gridded

data. For quantifying the similarity of simulated biome distributions and pollen-based biome reconstructions, we therefore

implement a new metric following both methods called the best neighbour score (BNS), accounting for agreement in the

235 neighbourhood of the record site and therewith being more tolerant for the position of the site. Within this metric, not only

the grid-box locating the record sites are used for comparison with the records, but also the surrounding grid-boxes (3 grid-

boxes in each direction for T31, 6 for T63, 1 for 10°grid). Similar to the Fuzzy Kappa statistic, the similarity in the

neighbouring grid-cells is expressed by a distance decay function. We here choose a Gaussian function (Eq.(4)), giving grid-

cells directly at the site proportional larger influence than grid-boxes far away.

240 $\quad w=e^{\frac{-1}{2}*\left(\frac{distance}{3}\right)^2}$ $\qquad$ with $\quad distance=\sqrt{dlon^2+dlat^2}$ (4)

The best neighbour is defined as the nearest grid-box within the neighbourhood agreeing with the reconstructed biome type.

The agreement for each record is then given by the distance weight (w) of the best neighbour in each neighbourhood. It is

equal to 1 if the grid-box locating the site indicate the same biome as reconstructed and it is equal to 0 if all grid-cells in the

neighbourhood disagree with the record. The BNS is the mean of all individual neighbourhood scores. For instance, a BNS

245 of approx. 0.82 or 0.46 means that the best neighbour grid-cell is among the grid-cell 'circle' next to the site-locating grid-cell

in T63 or T31, respectively. Accordingly, a BNS of 0.04 indicates a distance between the best neighbour and the site-locating

grid-cell of of 7.5° on a Gaussian grid. In contrast to the Fuzzy Kappa statistic, the BNS neither takes agreement by chance into account nor considers potential spatio-temporal autocorrelation.

3 Results

3.1 Comparison of the PFT-based and climate-based biome-distributions for the pre-industrial time-slice

For the pre-industrial time-slice, the PFT coverage of eight different Earth System Model simulations has been converted into mega-biome distributions (Fig.3). Additionally, the underlying pre-industrial climate states are used as forcing for the BIOME1 model (i.e. the classical way of biomisation) to calculate the mega-biome distributions in equilibrium with the simulated climate states (Fig. 4). Overall, the PFT-based biome maps look similar to the climate-based ones. All major biome belts can be reproduced using the new method, independent of the resolution or the complexity of the vegetation models. The biomisation based on the PFT coverage generally assigns more grid-cells to forest or woody biomes (e.g. savanna instead of grassland or desert) than the classical method. This is most noticeable in South America, where the area covered by tropical forests is strongly increased in the PFT-based biome-distribution (Tab.D1), being more in line with observations. Likewise, the savanna and/or forest biomes are more spread on the African continent for nearly all biomisations with the exception of CLIMBER and IPSL-ESM-T63.

The Asian forest regions are slightly larger in most PFT-based biome-distributions compared to the climate-based ones. This impression is reinforced by the fact that for CLIM-LPJ and IPSL-ESM-T63 the PFT-method suggests a pronounced boreal forest belt in Northern Asia, not only reducing the size of the grassland but also of the temperate forest area.

For North America, the PFT-based approach yields less forest for MPI-ESM-T63 and IPSL-ESM-T31 than shown by the climate-based biomisation. As a consequence, the North American prairie fits better to observations for the PFT-based biome distributions. In Alaska and north-western Canada, parts of the tundra regions suggested by BIOME1 tend to be replaced by boreal forest when using the new approach, which is generally more consistent to the observed vegetation.

The differences between the PFT-based and climate-based biome distributions can be caused by deficiencies in the biomisation methods, biases related to the imperfect vegetation models, or biases in the simulated climate. While the effect of shortcomings in the vegetation models can not be disentangled, the caveats of the PFT-based method and the effect of climate biases on the PFT-to-biome conversion are further discussed in Section 4.

3.2 Quantitative comparison of the PFT-based biome-distributions with reference biome maps

To quantify the skill of the new method to represent the global biome distribution, we compare the resulting biome maps with the modern potential natural vegetation cover estimated by Ramankutty and Foley (1999, RF99 in the following) and the pre-industrial biome reconstructions provided by the Biome 6000 project (Harrison, 2017). As target for the skill, the

level of agreement between the BIOME1 derived biome-distribution of the observed climate (CRU-TS4, years 1901-1930, cf. Fig.2) and the references data-sets are taken, i.e. κ of 0.68 and FSS of 0.13 (Fig.5) with respect to RF99 and a κ of 0.46 and BNS of 0.73 with respect to the biome reconstructions.

280 The PFT-based biome distributions agree well with the references, independent of the model. The Kappa statistic shows an overall agreement to RF99 between 0.54 (for MIROC-ESM) and 0.79 (for MPI-ESM-T63) revealing a good- to very good agreement (Fig.5). Likewise, all models reach in total the level of good skill in the FSS metric (0.01-0.27). This agreement is in line with or even better than the match between RF99 and the climate-based biomisation and between RF99 and the CRU-TS4 biomisation that is taken as target skill (Fig.5). Though, the spread between the individual models is larger for the PFT-

285 based method than for the classical approach.

As expected, the Kappa statistic indicates that the PFT-based biome maps compare worse with the reconstructions than with RF99, underestimating the similarity to the point reconstructions. κ ranges from 0.2 (poor) to 0.49 (fair), which is in line with the target skill and the metrics for the climate-based biome maps. The BNS additionally considering accordance in the neighbouring grid-cells of the record sites reveals a good to very good agreement of the PFT-based biome distributions and

290 the records (between 0.40 and 0.74), not much lower than the target skill and in accordance with the climate-based biomisations. For the MPI-ESM, IPSL-ESM-T31, HadGEM2-ESM and CLIMBER biomisation, the PFT-based method even produces biome distributions that fit better to the biome reconstructions than the climate-based biome maps.

Despite the overall agreement between the PFT-based biome-distributions and the references, a closer look at the representation of individual mega-biomes in the converted maps indicate large differences among the models as well as

295 among the individual mega-biomes (Fig.6). While tropical forests and deserts compares best with RF99, the biome warm-temperate forest is not reproduced, independent of the underlying model simulation.

The skill for simulating the other biomes is very different for the diverse models. κ spreads from poor for one model to very good for other models. Correcting the PFT distribution in land-use areas by redistributing the area fraction to the other tiles has no impact on the performance of the method. The biome maps based on simulations applying land-use do not compare

300 worse with RF99 than the maps of other simulations. Likewise, the complexity of the vegetation model and the number of distinguished PFTs has no significant effect on the representation of the biome distribution, indicating that the climate limits used in the biomisation procedure are appropriate for the assignment of the PFTs to the distinct PFT-groups. The differentiation of the PFT types (e.g. the different forest types) in vegetation models is often based on similar climate limits, regardless of whether the model is a complex dynamic vegetation model, or a simple biome model. With the exception of the

305 PFT-based biomisation for CLIMBER, in which the coarse grid is clearly disadvantageous for capturing the reconstructed desert belts and the rather regionally confined biomes (savanna, warm-temperate forest), the spatial resolution of the models is not the primary factor for the spread in the metrics. The PFT-based method performs equally well for simulations using T63 as for simulations using T31 (in total), only the regionally distributed warm-temperate forest is better represented in finer resolutions.

In general, the skill in representing the individual mega-biomes is similar for the PFT- and the climate-based method. Both approaches have the same strengths and weaknesses, but the spread between the models is larger for the PFT-based biomisations. In comparison with the climate-based method, the tropical, the warm-temperate and the boreal forest biomes tend to be slightly better represented by the PFT-based method. In contrast, the temperate forest, savanna and grassland distribution - averaged over all models – fit better to RF99 when using the climate-based approach, although for individual simulations, κ derived for the PFT-biomisation exceeds the climate-based one. The savanna and grassland biomes are particularly misrepresented in the biome maps that are based on the PFT-distributions simulated by MIROC-ESM, CLIMBER, CLIM-LPJ and HadGEM2-ESM. The temperate forest is poorly reproduced only in the PFT-based biomisation of CLIM-LPJ. Overall, the metrics indicate that the PFT-based method works similarly well as the classical approach of biomising climate states via the BIOME1 model. Likewise the method is able to keep up with the method by Prentice et al. (2011), further discussed in the Appendix E.

3.3 PFT-based biome-distributions for the mid-Holocene and Last Glacial Maximum time-slice

The sensitivity of the PFT-based method to changes in the vegetation cover is assessed by evaluating palaeo-biome distributions. For the mid-Holocene time-slice, four different simulations have been analysed. The main vegetation change described by biome reconstructions are the northern shift of the northern hemisphere forest-belts, in particular a northward displacement of the taiga-tundra boundary, and the decrease of the desert areas compared to pre-industrial. According to the BIOME6000 records, grassy vegetation reached at least up to 26°N at 6k, far into the modern central Sahara (Fig.7). For none of the models, this biome shift is reproduced, neither in the PFT-based (Fig.7) nor in the climate-based biome distributions (not shown). The mean Sahara desert border shifts northward by one to two grid-cells in the biomisations (i.e. ca. 1.875° to 3.75°, Tab.4). This shift collocates with substantial reductions in the desert fractions simulated by the individual ESMs (Fig.8). Only in MPI-ESM-T63, vegetation is increased in the entire western and central Sahara, but this increase is lower than 20%, not leading to a change in the biome assignment from desert to grassland. As the climate-based biomisations performed with BIOME1 reveal a reduction of the Sahara desert area in the same magnitude as the PFT-based ones, we conclude that the new biomisation method shows a reasonable sensitivity to the simulated changes in the desert fractions.

For all models with the exception of MIROC-ESM, the PFT-based biomisation reproduces an increased forest biome fraction in Eurasia north of 60°N during the mid-Holocene compared to pre-industrial, in line with the biome reconstructions (Tab.5). Though the magnitude of the change differs between the models, ranging from 0% within MIROC-ESM to 12% within CLIM-LPJ. For nearly all models (except for MIROC-ESM), the expansion of the forested area in the high northern latitudes seen in the PFT biomisation is of similar magnitude as in the climate-based biomisation, confirming that the method covers past vegetation changes with reasonable sensitivity.

Overall, the biome distributions for the mid-Holocene compare equally well to the reconstruction as for the pre-industrial time-slice (Fig.9). Although, κ is similarly low, ranging from 0.17 for the CLIM-LPJ biomisation to 0.38 for the MPI-ESM-T63 biomisation (poor agreement), the spread in the models and the differences in κ between the PFT-based biomisations and the climate-based biomisations are nearly identical to the results for the pre-industrial biome distributions.

345 In line with the results for PI, the BNS indicate a good to very good agreement to the biome reconstructions (ranging from 0.44 for MIROC-ESM to 0.72 for MPI-ESM-T63). The skill to capture the reconstructed individual mega-biomes strongly depends on the number of available pollen records, thus, temperate and boreal forests are represented best (Fig.7), while the simulated savanna regions are not supported by the biome reconstructions.

For the Last Glacial Maximum time-slice, five different simulations have been analysed. According to BIOME6000, the
350 main reconstructed vegetation differences at LGM compared to PI are a strong equator-ward retreat of the forest biomes and an expansion of tundra and steppe regions. The northernmost record indicating boreal forest during LGM is located at approx. 51°N in Asia (Fig.10). The PFT-biomisations mostly reproduce this reduction and the shift in Northern Hemisphere forest biomes (Fig.10), though the extent of the shift is underestimated. Forest reaches up to 50°N (for CLIMBER) to 65°N (for MPI-ESM-T63). The boreal forest position in the MIROC-ESM biomisation is not much changed compared to PI, but
355 boreal forest nearly replace the temperate forest biome.

The overall agreement of the PFT-based biome distributions with the biome reconstructions is rather fair, but in line with the results for the climate-based biome distributions. κ ranges from 0.07 for MPI-ESM-T31 to 0.23 for CLIMBER, only indicating a poor similarity of the biome maps and records (Fig.11). The BNS ranges from 0.24 (for MIROC-ESM) to 0.57 (for MPI-ESM-T63) revealing a fair to good agreement. The values for both metrics are in the same magnitude as for the
360 climate-based biomisations. Similar as for the PI time-slice, neither the complexity nor the spatial resolution are the main reason for the differences between the PFT-biomisations. The spread in the skill of representing the individual biomes is large, and no systematic bias for one model can be found. With the exception of the biomisation for CLIMBER, the savanna biome is misrepresented in all biomisations, independent of whether the PFT-based or the climate-based method was used. Within the model ensemble, tropical and temperate forest can be reproduced best.

365 **4. Discussion**

4.1 Caveats in the method

Even if the biomisation is restricted to mega-biome level, no clear definitions exist to distinguish biomes in terms of plant functional type compositions. While the bioclimatic limits used in the biome models are based on empirical analysis, no equivalent classification regulates the biomisation of PFTs. We particularly face this problem in finding a meaningful
370 threshold of maximum tree cover needed for defining forests. When is an accumulation of trees identified as forest? As models tend to underestimate the forest coverage and forest extent in the high northern latitudes (cf. Loranty et al., 2013) we

choose the assumption of tree cover being just dominant in forested grid-cells, although this limit is very low. We test other limits (e.g. absolute dominance, i.e. fractional coverage exceeding 50%), but these work worse, for most simulations used in this study as well as other simulations.

The mega-biome 'warm-temperate forest' (e.g. subtropical forest) includes PFTs that can be assigned to several biomes and is rather defined by a co-existence of certain PFTs. For instance, in the BIOME4 model it is not only defined by the dominance of temperate evergreen broadleaved trees, but can also be defined by a dominance of cool conifers (with sub-PFT of temperate evergreen broadleaved trees). The cool conifers – in turn - are also part of temperate forest biomes. Given the limited number of PFTs in the DGVMs , the confinement of biomes via PFT-mixtures is not possible. As biome models such as BIOME4 generally manage to simulate warm-temperate forest at the correct locations, we adopt the bioclimatic limits from BIOME4 (limit for temperate evergreen broadleaved trees) for defining this mega-biome. Nevertheless, the calculated warm-temperate forest distribution strongly disagree with the reference datasets. The warm-temperate forest shares most subtropical tree species with the tropical evergreen forest (Ni et al, 2010). These biomes tend to overlap in some regions and are sometimes mixed up in reconstructions (Chen et al, 2010). In addition, this mega-biome includes the 'warm temperate rainforest' and the 'wet sclerophyll forest and woodland' in the BIOME6000 reconstructions (cf. Harrison, 2017), which may not be able to be identified with our biomisation method. Regarding the modern reference of RF99, we decided to assigned the biome 'temperate needle-leaf evergreen forest and woodland' of the RF99 dataset to the mega-biome 'temperate forest', although, this biome is also located e.g. in the southern USA which should be assigned to the warm-temperate forest. Therefore, the evaluation of this method with respect to warm-temperate forest might be ambiguous. Furthermore, warm-temperate forests are small and rather patchily distributed and are thus rarely dominant in the coarse grid-cells to which the RF99 reference had to be interpolated to. The coarser the grid is the more warm-temperate forest regions get lost during the interpolation. Therefore, the warm-temperate-forest biome is generally better represented for models using a higher spatial resolution (i.e. MPI-ESM-T63, CLIM-LPJ and IPSL-ESM-T63).

In addition, not all biomes can be differentiated by the structural composition or climatic tolerance. The biome 'savanna' is the second-largest ecosystem in the tropics, covering approx. one fifth of the global land surface (Scholes and Hall, 1996). It occurs in climatic zones that are also suitable for forest and grasslands (Lehmann et al., 2011) and is, thus, very variable regarding the plant composition. Tree fraction can vary from very dense (open forest savanna) to nearly zero (Torello-Raventos et al., 2013). While tropical savannas require the coexistence of trees and C4-grass, they can only be distinguished from forests by their unique functional ecology, fire tolerance and shade intolerance (Ratman et al., 2011). These features make savannas unstable and vulnerable to changes in e.g. grazing, fire regime and climate, transforming savannas into forest or grasslands (Franco et al., 2014). The functional diversity of savannas is not adequately included in DGVMs nor considered in the biomisation method presented here. As even C4-grass is not simulated in all models, we had to define the savanna biome in a very rudimentary manner by a mixture of woody PFTs and grass and by bioclimatic limits, i.e. a mean temperature of the coldest month exceeding 10° which is taken as limit for C4-grass in dynamic vegetation models (e.g. JSBACH; cf. Reick et al., 2013). The savanna biome might therefore not be represented well. At least in the palaeo-

25   13

simulations, most biomisations do not capture the reconstructed savanna area, but this may partly also be related to the fact that only few records exist indicating savanna during LGM and 6k. Modern pollen rain analysis reveal that woody plant taxa typically characterizing Savannas are under-represented or even absent in the pollen/vegetation ratios. Given the lack of savanna indicators, this biome may be overlooked in fossil pollen records (Jones et al. 2011).

410 Similar to dynamic vegetation models, the priority in the biomisation procedure is given to forest biomes. It is first tested, whether forest biomes are suitable for covering the grid-cell, before savanna is distributed. Grasslands and tundra are assigned to the residual grid-cells, independent of the real grassy PFT cover fractions. The only restriction is a total vegetation coverage exceeding 10% for tundra or 20% for grassland to be distinguishable from deserts. This method has the large disadvantage that biases in the forest distribution propagate throughout the assignment of all biomes with the exception

415 of deserts. The forest biome distribution calculated for the different models is further tested in Section 4.3 for the pre-industrial time-slice.

Another problem is the inclusion of anthropogenic plant functional types (i.e. land-use) in some simulations, making the biome distribution less comparable to the reference data. Although land-use is often prescribed in the models, this process cannot be reversed in the final output data. The area chosen for land-use is historically determined and is based on human

420 decisions and not primarily on climate conditions. These human pathways cannot be reproduced in simple biomisation methods nor in the current dynamic vegetation models. We artificially rescale the natural vegetation in human-affected regions by redistributing the fraction of anthropogenic PFT coverage proportionally to the natural PFTs. This is a very simple approach and only partly in line with the implementation of land-use in the dynamic vegetation models. For instance, within JSBACH pasture is preferentially assigned to natural grasslands, forests are only affected if prescribed pasture fraction

425 exceeds the natural grassland area (cf. Reick et al., 2013). This rule is plausible, but not reversible and therefore not appropriate for the biomisation method presented here. The results show that biome maps based on models including land-use do not agree worse with the references than the other simulations, underlining that the redistribution method used here provides a good approximation of the natural vegetation cover.

The method of PFT-biomisation basically uses bioclimatic limits that have been inferred for the modern vegetation-climate

430 relationships. These limits may not be valid for all time periods. During LGM, the atmospheric CO_2-concentration was substantially lower than today, which may also affect the response of the plants to the background climate. Likewise, the matrices for the assignment of taxa into PFTs and from PFTs into Biomes have been constructed on the basis of the recent vegetation. These matrices do not have to be constant in time, i.e. they may not be applicable for glacial vegetation. Furthermore, the classification of the biomes itself corresponds to the modern vegetation and does not necessarily have to

435 reflect the palaeo-vegetation. There might be other biomes in glacial climates that remain unconsidered. This may lead to biases in the modelling results and the biome reconstructions taken as reference. The κ values for the LGM time-slice were quite low, indicating disagreement between the simulated and reconstructed biome distributions.

A rather technical problem is the interpolation of the PFT-distributions to the T31 or T63 grid that partly leads to a decrease in the global area to be compared with the reference datasets due to a mismatch of the land-sea-masks. In regions with a

440   strong change of the PFT fractional composition (e.g. desert border, coastal region), the interpolation may produce blurry transitions in the PFT distributions resulting in an erroneous conversion into the mega-biomes.

4.2 Biases in the pre-industrial biome distributions and the influence of the climate background

The classical method of biomising climate states via biome models (here BIOME1 by Prentice et al., 1992) and the new PFT-based method result in similar biome-distributions for most models and all time-slices. Generally, the PFT-based
445   method produces more forest in comparison with the classical approach. This is mainly related to the rather low limit of forest fraction needed in the assignment of forest in the PFT-biomisation procedure. In forest regions, where Earth System Models tend to produce large biases in the climate state, the PFT-based approach may be more suitable for the biomisation. Therefore, the tropical, warm-temperate and boreal forest are probably better represented by the PFT-method. However, the biomisation of the PFT-distributions itself strongly depends on the underlying climate, affecting both the differentiation into
450   the biomes as well as the simulation of the PFT coverage in the different dynamic vegetation models. To accurately compare the performance and the skill of the different vegetation models to represent biome distributions, the models should therefore be forced by the same climate state, but only few models can be run offline. This study is thus not thought as model evaluation, but as introduction to the biomisation method and as test whether the procedure works for models of different complexity and simulations for different time-slices.

455   To assess the contribution of the effect of biases in the underlying climate to the differences in the PFT-based biomisations and the references, we compare the pre-industrial climate-based biomisations with the biomisation of the CRU-TS4 dataset. A sensitivity study is performed following Dallmeyer et al. (2017) to relate differences in the biome distributions to precipitation or temperature deviations in the background climate (Fig.12). For this purpose, we successively replace the temperature or the precipitation in the CRU-TS4 forcing file for the BIOME1 model with the respective pre-industrial
460   temperature or precipitation distributions simulated by the models. Afterwards, we compare the differences between the calculated biome distributions.

Generally, disagreement in the high northern latitudes is associated with biases in temperature while disagreement in low latitudes co-occurs with precipitation biases for the pre-industrial time-slice. The similarity of the PFT-biomisations and RF99 is lowest for CLIMBER, MIROC-ESM and CLIM-LPJ. While for CLIMBER the coarse resolution (i.e. the very
465   different land-sea masks) may be the main responsible factor for disagreement, total κ is reduced by an underestimation of grasslands and savanna and an overestimation of the forests in the PFT-based biomisation of MIROC-ESM. This is exactly opposite to the biases occurring in the climate-based biomisation, indicating that the climate is not the primary origin of the differences. For this specific model, the PFT-based biomisation strongly differs from the climate-based one. This may at least partly be related to the handling of vegetation in the model. The vegetation model 'SEIB' included in MIROC-ESM is a
470   forest gap model, not using the tiling approach (Sato, 2007). PFT fractions have only been estimated during the CMIP5 post-processing, based on the net-primary-productivity ratios of the different vegetation categories. This approach might lead to

an overestimation of forest. On the other hand, according to BIOME1 the underestimation of the tropical forest domain in the climate-based biomisation of MIROC-ESM is caused by the way to dry climate in South America.

In the PFT-biomisation of CLIM-LPJ, κ is basically reduced by an overestimation of boreal forest at the expense of
475 temperate forest and an underestimation of the savanna regions. Both errors are mainly not climate driven. Within dynamic vegetation models explicitly calculating boreal and temperate forest, these forest types can coexist. To give a clear assignment, we decided to differentiate both forest types by the dominant tree PFT, i.e. if the boreal tree fraction exceeds temperate tree fraction, forest fraction is assigned to boreal forest and vice versa. This partly disagrees with the handling in biome models, as e.g. in cool mixed forest, boreal trees could be the dominant PFT and temperate trees only the sub-
480 dominant PFT (c.f. Kaplan et al., 2003), but this biome would be assigned to the mega-biome 'temperate forest'. We assume that due to a slight overestimation of boreal forest coverage in Europe and at the modern boreal to temperate forest transition zone within CLIM-LPJ, the vegetation in these regions is grouped into the mega-biome 'boreal forest'. In South America, tropical forest fraction is overestimated by CLIM-LPJ with values exceeding 80% in most regions of Brazil, precluding the savanna biome. Within North Africa, CLIM-LPJ simulates hardly any regions with coexisting substantial forest and grass
485 fractions. Either tropical trees are clearly the dominant PFT (assigned to tropical forest) or forest fraction is too low (below 10%) to be assigned to savanna. The defined limits for savanna are only fulfilled for very few grid-cells.

For MPI-ESM-T31, the boreal forest biome is strongly underestimated in the PFT-based biomisation. The BIOME1 results clearly relate this bias to a too cold climate (GDD5 limit is not reached in BIOME1) which also affects the simulation of trees in JSBACH sharing the same bioclimate limits. Therefore, forest fraction in MPI-ESM-T31 is underestimated for the
490 northern latitudes.

IPSL-ESM-T31 shows a dry bias in South America resulting in a too low tropical forest biome cover, in both, the climate-based and the PFT-based biomisation. BIOME1 reveals another systematic bias for the MIROC-ESM biomisation indicating too much temperate forest in North America at the expense of grassland and partly of boreal forests. This overestimation of temperate forest is induced by a too wet climate favouring growing of trees and a rather too warm climate in the high
495 northern latitudes.

4.3 Evaluating the distribution of forest biomes

Due to the forest priority rule in the biomisation method, the skill of representing the non-forested biomes depends on how well the forest distribution can be reproduced. To further assess the performance of the method with respect to the forest biomisation, we analyse the pre-industrial zonal mean forest fraction in form of the zonal sum of forested area per latitude to
500 be independent of the different grid sizes used for the individual simulations (Fig.13). In nearly all PFT-biomisations, the zonal forest fraction is underestimated in the high northern latitudes and the zonal maximum is shifted southward, although the defined limit of minimum required tree fraction is already quite low in the PFT-method. This bias is most obvious in the biomisation for MPI-ESM-T31 and IPSL-ESM-T63. While for MPI-ESM-T31, the coexistence of the bias in both, the

climate-based and the PFT-based biomisation, underlines the effect of the too cold climate on the forest distribution, the strongly shifted high latitude forest maximum in the IPSL-ESM-T63 PFT-biomisation is probably not climate driven.

In the tropical regions, the forest fraction tends to be underestimated when using the climate-based method whereas forest fraction is often too high in the PFT-based biomisations, probably related with the low tree fraction limit needed for forest assignment in the PFT-method. The tropical forest fraction based on the simulated climate and PFT-distribution by CLIMBER is strongly overestimated. This is at least partly caused by the coarse grid and specific land-sea-mask used in the model.

To further quantify the biases in forest fraction, we compare the centred root mean square error (cRMSE), the Pearson correlation coefficient (r) and the zonal variability between the biomisations of the ESM simulations and the RF99 reference, combined in a Taylor-diagram (Fig.14). Overall, the zonal forest fraction in the biomisations agrees well with the reference. All biomisations show a good to nearly perfect pattern correlation with values exceeding 0.77, independent of the chosen method. For most models, the Pearson correlation coefficient even exceeds 0.9. The PFT-based biomisation is worst for CLIMBER and MIROC-ESM, revealing a too large standard deviation and a cRSME of 0.83 and 0.73, respectively. For MPI-ESM-T31, spatial variability is slightly too low and the cRSME is 0.61 using the PFT-based method and 0.63 in the climate-based biomisation, reflecting the common underestimation of the boreal forest. As expected, best performance can be observed for the biomisations based on simulations with prescribed PFT coverage (MPI-ESM-T63 and IPSL-ESM-T31), sharing a similar standard deviation with RF99, a pattern correlation coefficient of 0.98 and a cRMSE of 0.21, which is even better as the biomisation of the CRU TS4 data (cRMSE of 0.28).

5 Summary and Conclusion

Dynamic global vegetation models use different kinds and numbers of plant functional types to represent the global vegetation. These PFT distributions can neither be directly compared between different models nor between models and reconstructions, which were hitherto mostly provided in form of biomes. We therefore have developed a method for biomising simulated PFT distributions and have tested this method for six state-of-the art dynamic global vegetation models based on simulations for the pre-industrial, mid-Holocene and Last Glacial Maximum time-slices.

Overall, the method works well for all models and can keep up with other biomisation techniques. The comparison with different references datasets (i.e. pollen-based biome reconstructions and estimates of the potential natural vegetation) reveals a similar agreement with the PFT-based biomisation than with biome distributions inferred from the biome model BIOME1 (Prentice, 1992) that has been forced with the background climates. The comparable skill to the BIOME1 model, which is tuned to represent the global vegetation as well as possible, is partly achieved by the use of bioclimatic limits that are in line with the definitions in biome models.

The skill of capturing the global biome distributions is independent to the spatial resolution and the complexity of the vegetation models or the integration of land use. For models just using two different PFTs (CLIMBER) the method performs

equally well as for models using ten different PFTs (e.g. IPSL-ESM). Only the very coarse resolution in the CLIMBER model hampers the comparability with the single point reconstructions, in particular for biomes with a very limited number of available records. In addition, the quantitative comparison of the biomised vegetation distributions among each other and with the gridded reference data is complicated by the very different model resolutions.

540   In general, large biome belts (such as tropical forest) can be captured best, while rather regionally confined biomes such as savanna and warm-temperate forest are not as well represented. This may at least partly be related to the fact that these biomes can not be defined clearly via PFT cover fractions. Savannas are characterized by a distinct functional ecology and can not be differentiated from other tropical biomes via plant composition or climatic tolerance. The warm-temperate forest is rather defined by a coexistence of PFTs an might overlap with other biomes. This may lead to mismatches with the

545   reconstructions. For the palaeo-simulations, the agreement between the individual mega-biome distributions derived by the PFT-method and the biome reconstructions strongly depends on the number of available records. The main vegetation differences between the pre-industrial and mid-Holocene or Last Glacial Maximum time-slices are captured by most models and are also reflected in the PFT- and climate-based biomisations, indicating a reasonable sensitivity of the PFT-method. In total, the Kappa statistic reveals only poor agreement between the PFT- and climate based biomisations and the

550   reconstructions for LGM, which might be related to the use of bioclimatic limits inferred from modern observations that may not be valid for climate states being totally different from present day's.

We have provided a simple but powerful method for the biomisation of simulated plant functional type distributions that requires only few input variables and can hence be applied to all kind of dynamic global vegetation models. The new method

555   can keep up with the classical biomisation approach of forcing biome models with climate states. However, as the new biomisation of the simulated PFT-fractions indirectly accounts for all processes included in the dynamic vegetation models (e.g. ecophysiological response of the plants to changes in the environment such as atmospheric CO_2-level), the new method is able to more directly represent the output of the vegetation modules of an Earth System Model. The biomisation of the simulated vegetation, thus, facilitates the direct comparison between different Earth System Models, and between models and biome

560   reconstructions. It is therefore a powerful method for the evaluation of Earth System Models, particularly suitable for the assessment of recent palaeo-vegetation changes.

6 Code and data availability

The PMIP3 simulations of MPI-ESM-T63, ISPL-ESM-T31, MIROC-ESM, and HadGEM2-ESM can be downloaded from

565   the Earth System Grid Federation. Simulation IDs are listed in Table 4.

35   18

The tool for the biomisation of PFT-distribution, input data, other scripts used in the analysis and supplementary information that may be useful in reproducing the authors' work will be archived by the Max Planck Institute for Meteorology and can be obtained by contacting publications@mpimet.mpg.de

Acknowledgements

570 This work contributes to the project PalMod, funded by the German Federal Ministry of Education and Research (BMBF), Research for Sustainability initiative (FONA, www.fona.de). AD was financed by PalMod. We thank J. Koch (GEUS) for fruitful discussion on metrics and T.Kleinen (MPI-M) for performing the CLIMBER and CLIM-LPJ simulations and his helpful comments on an earlier version of this manuscript. We furthermore thank Dan Zhu (IPSL) for providing the IPSL-ESM-T31-simulations and the biome data derived by the FPC-biomisation method. We would like to thank the anonymous

575 Referees for her/his constructive comments which helped to significantly improve the manuscript.

We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in Section 2.2 of this paper) for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for

580 Earth System Science Portals.

A Biomisation of the PFT-distribution using the example of VECODE (included in Climber2)

Fig. A1 shows the PFT-biomisation procedure based on the VECODE model (include in Climber2). VECODE includes only
585 2 PFTs, therefore all rules defined in the biomisation method are needed. For the other models using more PFTs, the flowchart would look slightly different.

B Simulations

The MPI-ESM-P (Giorgetta et al., 2013) simulations have been performed at the Max-Planck-Institute for Meteorology and include the land model JSBACH with dynamic vegetation module (c.f Reick et al., 2013). In the pre-industrial control
590 simulation, vegetation pattern and land use were prescribed. For the palaeo-simulations, we use the simulations with interactive vegetation. The spatial resolution for the atmosphere and land is T63 (i.e. approx. 1.875° on a Gaussian grid). These simulations are referred to as MPI-ESM-T63 in the following. In a similar model setup, additional PMIP3-like experiments have been undertaken for PI and LGM by Klockmann et al. (2016) in a coarser spatial resolution (T31, i.e. approx. 3.75° on a Gaussian grid, MPI-ESM-T31).
595 IPSL-CM5A-LR (Dufresne et al. 2013) is the low resolution CMIP5 model version of the Institute Pierre Simon Laplace and contains the terrestrial biosphere model ORCHIDEE (Krinner et al., 2005) that is run offline, forced with climate input. In the PMIP3-simulations, vegetation and land use were prescribed. For better comparison with the gridded reference dataset (see next section), the climate and PFT-fields have been interpolated bilinearly to a Gaussian T31 grid. Using the simulated LGM climate of the PMIP3 simulation, Zhu (2016) performed additional experiments for LGM with ORCHIDEE-MICT
600 (Guimberteau et al., 2018), a model version with improved vegetation dynamic in the high northern latitudes (Zhu et al., 2015). The corresponding PI control simulation has been forced by CRUNCEP v.5.3.2 data (http://dods.extra.cea.fr/store/p529viov/cruncep/V4_1901_2012/readme.htm), that is a combination of observations (CRU data) and reanalysis data (NCEP). The simulations have been interpolated to a Gaussian T63 grid.
HadGEM2-ESM (Collins et al., 2011) is the Earth System Model of the Met Office Hadley Centre and includes the
605 vegetation model TRIFFID (Cox, 2001). In all simulations used here, the model ran with interactive vegetation. The PI simulation (piControl) include land-use types. The simulations have been remapped to a Gaussian T63 grid.
The dynamic vegetation model LPJ (Lund-Potsdam-Jena model, Stich et al., 2003) is usually used for offline simulations, forced by climate simulations or observations. The simulations used here has been conducted in a similar model-setup as described in Kleinen et. al. (2010), but has been re-done on a new computer (T. Kleinen, personal communication), which
610 may lead to very small deviations from the original runs. The PI simulation has been forced by observational datasets (CRU

TS3.1, Harris et al., 2014), the 6k simulation by output from the CLIMBER-2 model. Both simulations have been interpolated to a Gaussian T63 grid and are referred to as CLIM-LPJ in the following text.

MIROC-ESM (Watanabe et al., 2011) is the Earth System Model of the Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental
615 Studies. It includes the dynamic vegetation model SEIB (Sato et al.,2007). SEIB deviates from the other DGVMs in this study as it does not use the tiling approach of calculating PFT fractional coverage for each grid-cell. It is a so called gap-model, simulating the interactions among individual trees that compete for light and space in arising gaps (e.g. due to disturbances) within a spatially explicit virtual forest. The model was built for capturing the vegetation dynamics on local scale. The application of the model for larger (e.g. global) scales is possible, but global simulations partly disagreed with
620 observations (Sato et al., 2007). The PFT distribution used in this study has been calculated in the post-processing for CMIP5 via the relative net primary productivity of the vegetation categories, it was not explicitly calculated by the model, which may lead to additional biases in the vegetation distribution. The simulation for PI (piControl) includes land use. These simulations have been remapped to a Gaussian T31 grid.

CLIMBER2 (Petoukhov et al., 2000) is an Earth System Model of intermediate complexity and contains the vegetation
625 module VECODE (Brovkin et al., 1997). The LGM and PI simulations have been specifically undertaken for this study (T. Kleinen, personal communication) and are referred to as CLIMBER in the following. The CLIMBER output has not been interpolated as the simulation ran with a too coarse resolution of 10° latitude x 51° longitude. To compare with the data and the other models, the CLIMBER output was re-gridded to 10°x10° grid without interpolation.

The dynamic global vegetation model 'CLM-DGVM' as part of the Community Earth System Model (Hurrell et al., 2013) is
630 currently under re-development. No appropriate simulations could be provided.

C Preparing the reference datasets

Most biomes used in RF99 could directly be assigned to the mega-biome types used in this study (Tab. C1). RF99 additionally includes the biome 'Evergreen/Deciduous Mixed Forest/Woodland'. Here, this biome is assigned to the mega-biomes 'temperate forest' in warm regions and 'boreal forest' in colder regions via the modern growing degree days
635 distribution (GDD5≥900°C for warm region, GDD5<900°C for cold region, cf. Tab.2), derived from observations (University of East Anglia Climatic Research Unit Time Series 3.1, University of East Anglia, 2008). Likewise the savanna biome had to be split up as RF99 includes temperate savanna which is explicitly excluded in the definition of the savanna mega-biome used in this study. The threshold for warm savanna is a mean temperature of the coldest month exceeding 10°C (Limit for existence of C4 grass in JSBACH). To compare RF99 with the different model simulations, RF99 had to be
640 remapped to the model grids. We decided to use the spatial resolutions T31, T63 and also prepared a map for the downscaled CLIMBER output (10°x10°grid). Within each of this model grid cells, the dominant mega-biome type in the 5-minute-resolved RF99 data was taken for covering the RF99 grid-box in T31 or T63 or in the 10°grid. In more detail, each grid-box on a T31 Gaussian grid contains 45*45 grid-cells of the 5-minute-resolved RF99 data. Within these 45*45 grid-boxes the

fractional coverage of all mega-biomes is calculated and the biome with the highest fraction is chosen for covering the T31
645 grid-box. For T63 arithmetically 22.5*22.5 grid-cells form one T63 grid-cell. Here, we take 23*23 RF99 grid-boxes with one grid-box overlap to equally distribute the 5 minute grid-cells to the T63 grid. We are aware of the fact that the latitudes in the Gaussian grids are actually not equidistant, so that the remapped RF99 biome distributions are slightly stretched towards the poles, but this effect is marginal and is not expected to shift the main biome belts. To compare equal number of grid-cells, the reference data is cut by the land-sea masks used in the individual simulations to only include grid-cells being on land in both
650 data-sets.

E Comparison of the biomisation method with the approach of Prentice et al.

Prentice et al. (2011) introduced a biomisation method (further referred to as FPC method) that is fundamentally different to the method presented in this study. The assignment to the different biomes is controlled by the foliage projective cover, the
655 vegetation height and the PFT composition. Climatic limitations in form of growing degree days are only used to distinguish the tundra biomes.

Unfortunately, the foliage projective cover and the vegetation height are not included in the standard output of the vegetation models and are therefore not available for the simulations used here.

To compare both methods, we therefore only use the simulations performed within IPSL-ESM-T63, which has been
660 biomised by Zhu (2016) following the approach of Prentice et. al. (2011). The biome output has been grouped into mega-biomes and remapped to a T63grid in the same way as the RF99 reference were prepared.

Beside the savanna regions, the derived biome map resembles the map resulting from the PFT-based method (Fig. E1). Prentice et al. distinguish temperate parkland, sclerophyll woodland and boreal parkland that all have been assigned in the mega-biome savanna, but savanna is only defined as tropical savanna in our method. This complicates the comparison of the
665 biome maps and leads to strong differences in the savanna distribution between both methods and between the biomisation using the FPC method and RF99, although we leave the temperate savanna in RF99 for better comparison.

Since boreal parkland is not included in RF99, the PFT-method introduced here yield better results for boreal forest and savanna than the FPC method. Additionally, the warm-temperate forest is more appropriately reproduced. In contrast, temp. forest and grassland is better represented using the FPC method. All other biomes are equally well simulated for the PI time-
670 slice.

The FSS metric (Tab.E1) indicate that the PFT-based method (0.13) agrees in total slightly better with the reference than the other approach (0.10). Overall, the biomisation using the FPC method reaches a κ of 0.59 (vs. 0.63 for the PFT-based method) compared to RF99 and 0.19 (vs. 0.24) compared to the Biome6000 pollen data. BNS is 0.53 for the FPC- method and 0.55 for the PFT-method.

675   The LGM biome distribution can be captured slightly better using the FPC method (k=0.17 vs 0.13, and BNS=0.54 vs. 0.41). Particularly the tropical forest, desert and savanna biome agree better with the biome reconstructions than for the PFT-based method, but at least for the latter biomes, the record density is very low which may distort the results.

References

Bigelow, N.H., Brubaker, L.B., Edwards, M.E., Harrison, S.P., Prentice, I.C., Anderson, P.M., Andreev, A.A., Bartlein, P.J.,
680 Christensen, T.R., Cramer, W., Kaplan, J.O., Lozhkin, A.V., Matveyeva, N.V., Murray, D.F., McGuire, A.D., Razzhivin, A.Y., Ritchie, J.C., Smith, B., Walker, D.A., Galjewski, K., Wolf, V., Holmquist, B., Igarashi, Y., Kremenetskii, K., Paus, A., Pisaric, M.F.J., and Volkova, V.S.: Climate change and Arctic ecosystems—I. Vegetation changes north of 50°N between the last glacial maximum, mid-Holocene and present. J Geophys Res, vol 108. doi:*10.1029/2002JD002558*, 2003.

685 Braconnot, P., Harrison, S.P., Otto-Bliesner, B., Abe-Ouchi, A., Jungclaus, J., and Peterschmitt, J.Y.: The Paleoclimate Modeling Intercomparison Project contribution to CMIP5. CLIVAR Exchanges No 56 16:15–19, 2011.

Brovkin, V., Ganopolski, A., and Svirezhev, Y.: A continuous climate-vegetation classification for use in climate-biosphere studies. Ecological Modelling, 101, 251-261, 1997.
690

Calvo, M. and Prentice, I. C.: Effects of fire and CO_2 on biogeography and primary production in glacial and modern climates. New Phytol, 208: 987–994. doi:10.1111/nph.13485, 2015

Chen, Y., Ni, J., and Herzschuh, U.: Quantifying modern biomes based on surface pollen data in China. Glob. Planet. Chang.
695 74, 114-131, 2010.

Cohen, J.: A coefficient of agreement for nominal scales, Educ. Psychol. Meas., 20(1), 37–46, 1960.

Collins, W. J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., Hughes, J., Jones, C. D., Joshi, M.,
700 Liddicoat, S., Martin, G., O'Connor, F., Rae, J., Senior, C., Sitch, S., Totterdell, I., Wiltshire, A., and Woodward, S.: Development and evaluation of an Earth-System model – HadGEM2, Geosci. Model Dev., 4, 1051-1075, https://doi.org/10.5194/gmd-4-1051-2011, 2011.

Cox, P. M.: Description of the TRIFFID dynamic global vegetation model. Technical note 24, Hadley Centre, Met Office,
705 Exeter, UK, 2001.

Crucifix, M., Betts, R. A., and Hewitt, C. D.: Pre-industrial potential and Last Glacial Maximum global vegetation simulated with a coupled climate-biosphere model: diagnosis of bioclimatic relationships, Global Planet. Change, 45, 295–312, 2005.

710 Dallmeyer, A., Claussen, M., Ni, J., Cao, X., Wang, Y., Fischer, N., Pfeiffer, M., Jin, L., Khon, V., Wagner, S., Haberkorn, K., and Herzschuh, U.: Biome changes in Asia since the mid-Holocene − an analysis of different transient Earth system model simulations, Clim. Past, 13, 107-134, https://doi.org/10.5194/cp-13-107-2017, 2017.

Diffenbaugh, N.S., Sloan, L.C., Snyder, M.A., Bell, J.L., Kaplan, J., Shafer, S.L., and Bartlein, P.J.: Vegetation sensitivity to
715 global anthropogenic carbon dioxide emissions in a topographically complex region. Global Biogeochem. Cycles 17, 1067, doi:10.1029/2002GB001974, 2003.

Franco, A. C., Rossatto, D. R., de Carvalho Ramos Silva, L., and da Silva Ferreira, C.: Cerrado vegetation and global change: the role of functional types, resource availability and disturbance in regulating plant community responses to
720 rising CO2 levels and climate warming, Theor. Exp. Plant Physiol., 26, 19–38, https://doi.org/10.1007/s40626-014-0002-6, 2014.

Dufresne, J-L, Foujols, M-A, Denvil, S., Caubel, A., Marti, O., Aumont, O., Balkanski, Y., Bekki, S., Bellenger, H., Benshila, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Cheruy, F., Codron, F., Cozic, A., Cugnet, D.,
725 de Noblet, N., Duvel, J-P., Ethé, C., Fairhead, L., Fichefet, T., Flavoni, S., Friedlingstein, P., Grandpeix, J-Y., Guez, L., Guilyardi, E., Hauglustaine, D., Hourdin, F., Idelkadi, A., Ghattas, J., Joussaume, S., Kageyama, M., Krinner, G., Labetoulle, S., Lahellec, A., Lefebvre, M-P., Lefevre, F., Levy, C., Li, Z. X., Lloyd, J., Lott, F., Madec, G., Mancip, M., Marchand, M., Masson, S., Meurdesoif, Y., Mignot, J., Musat, I., Parouty, S., Polcher, J., Rio, C., Schulz, M., Swingedouw, D., Szopa, S., Talandier, C., Terray, P., Viovy, N., and Vuichard, N.: Climate change projections using the IPSL-CM5 Earth
730 System Model: from CMIP3 to CMIP5, Clim Dyn 40: 2123. *https://doi.org/10.1007/s00382-012-1636-1*, 2013.

Foody, G.M.: Status of land cover classification accuracy assessment. Remote Sens. Environ. 80, 185–201, 2002.

Franco, A. C., Rossatto, D. R., de Carvalho Ramos Silva, L., and da Silva Ferreira, C.: Cerrado vegetation and global
735 change: the role of functional types, resource availability and disturbance in regulating plant community responses to rising CO2 levels and climate warming, Theor. Exp. Plant Physiol., 26, 19–38, https://doi.org/10.1007/s40626-014-0002-6, 2014.

Gilleland, E., Ahijevych, D., Brown, B. G., Casati, B., and Ebert, E. E.: Intercomparison of spatial forecast verification
740 methods. *Wea. Forecasting*, **24**, 1416–1430, doi:https://doi.org/10.1175/2009WAF2222269.1, 2009.

Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Böttinger, M., Brovkin, V., Crueger, T., Esch, M., Fieg, K., Glushak, K., Gayler, V., Haak, H., Hollweg, H.-D., Ilyina, T.,Kinne, S., Kornblueh, L., Matei, D., Mauritsen, T.,

Mikolajewicz, U., Mueller, W., Notz, D., Pithan, F., Raddatz, T., Rast, S.,Redler, R., Roeckner, E., Schmidt, H., Schnur,

745   R., Segschneider, J., Six, K. D., Stockhause, M., Timmreck, C., Wegner, J.,Widmann, H., Wieners, K.-H., Claussen, M., Marotzke, J., and Stevens, B.: Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5, J. Adv. Model. Earth Syst., 5, 572–597, doi:10.1002/jame.20038, 2013.

Guimberteau, M., Zhu, D., Maignan, F., Huang, Y., Yue, C., Dantec-Nédélec, S., Ottlé, C., Jornet-Puig, A., Bastos, A.,

750   Laurent, P., Goll, D., Bowring, S., Chang, J., Guenet, B., Tifafi, M., Peng, S., Krinner, G., Ducharne, A., Wang, F., Wang, T., Wang, X., Wang, Y., Yin, Z., Lauerwald, R., Joetzjer, E., Qiu, C., Kim, H., and Ciais, P.: ORCHIDEE-MICT (v8.4.1), a land surface model for the high latitudes: model description and validation, Geosci. Model Dev., 11, 121-163, https://doi.org/10.5194/gmd-11-121-2018, 2018.

755   Hagen, A.: Fuzzy set approach to assessing similarity of categorical maps. Int. J. Geogr. Inf. Sci. 17, 235–249, 2003.

Hagen, A.: An improved Fuzzy Kappa statistic that accounts for spatial autocorrelation, Int. J. Geogr. Inform. Sci., 23(1), 61–73, 2009.

760   Handiani, D., Paul, A., and Dupont, L.: Tropical climate and vegetation changes during Heinrich Event 1: a model-data comparison, Clim. Past, 8, 37–57, doi:10.5194/cp-8-37-2012, 2012.

Handiani, D., Paul, A., Prange, M., Merkel, U., Dupont, L., and Zhang, X.: Tropical vegetation response to Heinrich Event 1 as simulated with the UVic ESCM and CCSM3, Clim. Past, 9, 1683-1696, https://doi.org/10.5194/cp-9-1683-2013, 2013.

765

Harris, I., Jones, P.D., Osborn, T.J. and Lister, D.H.: Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset. Int. J. Climatol., 34: 623–642. doi:10.1002/joc.3711, 2014.

Harrison, S.: BIOME 6000 DB classified plotfile version 1. University of Reading. Dataset.

770   http://dx.doi.org/10.17864/1947.99 , 2017.

Harrison, S. P. and Prentice, I. C.: Climate and CO2 controls on global vegetation distribution at the last glacial maximum: analysis based on palaeovegetation data, biome modelling and palaeo climate simulations, Glob. Change Biol., 9, 983–1004, 2003.

775 Harrison, S.P., Prentice, I.C., Barboni, D., Kohfeld, K., Ni, J., and Sutra, J.-P.: Ecophysiological and bioclimatic foundations for a global plant functional classification, Journal of Vegetation Science, 21: 300–317. doi:10.1111/j.1654-1103.2009.01144.x, 2010.

Harrison, S., Bartlein, P.J., Prentice, C.I.: What have we learnt from paleoclimate simulations? J. Quat. Sci., 31, pp. 363-385, 780 2016.

Haxeltine, A., and C.I. Prentice: BIOME3: An equilibrium terrestrial biosphere model based on ecophysiological constraints, resource availability, and competition among plant functional types, Global Biogeochem. Cycles, 10(4), 693-709, 1996.

785 Hurrell, J.W., Holland, M.M., Gent, P.R., Ghan, S., Kay, J.E., Kushner, P.J., Lamarque, J., Large, W.G., Lawrence, D., Lindsay, K., Lipscomb, W.H., Long, M.C., Mahowald, N., Marsh, D.R., Neale, R.B., Rasch, P., Vavrus, S., Vertenstein, M., Bader, D., Collins, W.D., Hack, J.J., Kiehl, J., and Marshall, S.: The Community Earth System Model: A Framework for Collaborative Research. Bull. Amer. Meteor. Soc., 94, 1339–1360, https://doi.org/10.1175/BAMS-D-12-00121.1, 2013

790 Jacobson Jr., G. L. and Bradshaw, R. H. W.: The selection of sites for paleovegetational studies, Quat. Res., 16, 80–96, 1981.

Jolly, D., Harrison, S.P., Damnati B., and Bonnefille, R.: Simulated climate and biomes of Africa during the Late Quaternary: comparison with pollen and lake status data. Quat. Sci. Rev.,17, 629-657,1998.

795 Jones, H. T., Mayle, F. E., Pennington, R. T., and Killeen, T. J.: Characterisation of Bolivian savanna ecosystems by their modern pollen rain and implications for fossil pollen records, Rev Palaeobot Palyno, 164, 223-237, https://doi.org/https://doi.org/10.1016/j.revpalbo.2011.01.001, 2011.

Kageyama, M., Braconnot, P., Bopp, L., Mariotti, V., Roy, T., Woillez, M-N., Caubel, A., Foujols, M.-A., Guilyardi, E., 800 Khodri, M., Lloyd, J., Lombard, and F., Marti, O.: Mid-Holocene and last glacial maximum climate simulations with the IPSL model: part II: model-data comparisons, Climate Dynamics, 40, 9-10, 2469, 2013.

Kaplan, J.O., Bigelow, N.H., Prentice, I.C., Harrison, S.P., Bartlein, P.J., Christensen, T.R., Cramer, W., Matveyeve, N.V., McGuire, A.D., Murray, D.F., Razzhivin, V.Y., Smith, B., Walker, D.A., Anderson, P.M., Andreev, A.A., Brubaker, L.B., 805 Edwards, M.E., and Lozhkin, A.V.: Climate change and arctic ecosystems: 2. Modeling, paleodata-model comparisons, and future projections. J. Geophys. Res., 108,8171. doi:10.1029/2002JD002559., 2003.

Kleinen, T., Brovkin, V., von Bloh, W., Archer, D. & Munhoven, G.: Holocene carbon cycle dynamics. Geophysical Research Letters, 37, doi:10.1029/2009GL041391, 2010.

Klockmann, M., Mikolajewicz, U., and Marotzke, J.: The effect of greenhouse gas concentrations and ice sheets on the glacial AMOC in a coupled climate model, Clim. Past, 12, 1829-1846, https://doi.org/10.5194/cp-12-1829-2016, 2016.

Koch, J., Mendiguren, G., Mariethoz, G., and Stisen, S.: Spatial sensitivity analysis of simulated land-surface patterns in a catchment model using a set of innovative spatial performance metrics, J. Hydrometeorol., 18, 1121–1142, https://doi.org/10.1175/JHM-D-16-0148.1, 2017.

Krinner, G., Viovy, N., de Noblet-Ducoudre, N., Ogee, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S., and Prentice, I. C.: A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system, Global Biogeochem. Cy., 19, Gb1015,10 doi:10.1029/2003gb002199, 2005.

Landis, J.R., Koch, G.G.: Application of hierarchical Kappa-type statistics in assessment of majority agreement among multiple observers. Biometrics 33, 363–374, 1977.

Lehmann, C. E. R., Archibald, S. A., Hoffmann, W. A., and Bond, W. J.: Deciphering the distribution of the savanna biome, New Phytol., 191, 197–209, https://doi.org/10.1111/j.1469-8137.2011.03689.x, 2011.

Loranty, M. M., Berner, L. T., Goetz, S. J., Jin, Y., and Randerson, J. T.: Vegetation controls on northern high latitude snow-albedo feedback: observations and CMIP5 model predictions, Glob. Change Biol., 20, 594–606, doi:10.1111/gcb.12391, 2013.

Mittermaier, M., Roberts, N., and Thompson, S. A.: A long-term assessment of precipitation forecast skill using the fractions skill score. Meteor. Appl., 20, 176–186, doi:https://doi.org/10.1002/met.296, 2013.

Monserud, R.A., and Leemans, R.: Comparing global vegetation maps with the Kappa statistic. Ecol. Model. 62, 275–293, 1992

Ni, J., Yu, G., Harrison, S.P., and Prentice, I.C.: Palaeovegetation in China during the late Quaternary: biome reconstructions based on a global scheme of plant functional types. Palaeogeogr Palaeoclimatol Palaeoecol 289:44–61, 2010.

Petoukhov, V., A. Ganopolski, V. Brovkin, M. Claussen, A. Eliseev, C. Kubatzki, and S. Rahmstorf, CLIMBER-2: A climate system model of intermediate complexity, Clim. Dyn., 16, 1–17, 2000.

845   Prentice, I.C., Cramer, W., Harrison, S.P., Leemans, R., Monserud, R.A., and Solomon, A.M.: A global biome model based on plant physiology and dominance, soil properties and climate. J. Biogeogr., 19, 117-134., 1992.

Prentice, I.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. Climate Dynamics 12, 185–194, 1996.

850

Prentice, I.C., Webb III, T.: BIOME 6000: reconstructing global mid-Holocene vegetation patterns from palaeoecological records. Journal of Biogeography 25, 997–1005, 1998.

Prentice, I.C., Jolly, D., and BIOME 6000 participants: Mid-Holocene and glacial maximum vegetation geography of the

855   northern continents and Africa. Journal of Biogeography 27, 507–519, 2000.

Prentice, I.C, Bondeau, A., Cramer, W., Harrison, S.P., Hickler, T., Lucht, W., Sitch, S., Smith, B., and Sykes, M.: Dynamic Global Vegetation Modeling: quantifying terrestrial ecosystem responses to large-scale environmental change. In: Canadell JG, Pataki DE, Pitelka LF, eds. Terrestrial ecosystems in a changing world. Berlin, Germany: Springer, 175-192, 2007.

860

Prentice, I.C., Harrison, S.P., and Bartlein, P.J. Global vegetation and terrestrial carbon cycle changes after the last ice age. New Phytol 189:988–998, 2011.

Ramankutty, N. and Foley, J. A.: Estimating historical changes in global land cover: croplands from 1700 to 1992, Global

865   Biogeochem. Cy., 13(4), 997–1027, 1999.

Ratnam, J., Bond, W. J., Fensham, R. J., Hoffmann, W. A., Archibald, S., Lehmann, C. E. R., Anderson, M. T., Higgins, S. I., and Sankaran, M.: When is a "forest" a savanna, and why does it matter?, Glob. Ecol. Biogeogr., 20, 653–660, https://doi.org/10.1111/j.1466-8238.2010.00634.x, 2011.

870

Reick, C. H., Raddatz, T., Brovkin, V., and Gayler, V.: Representation of natural and anthropogenic land cover change in MPI-ESM, J. Adv. Model. Earth Syst., 5, 459–482, 2013.

Roberts, N. M., and Lean, H. W.: Scale-selective verification of rainfall accumulations from high-resolution forecasts of

875   convective events. Mon. Wea. Rev., 136, 78–97, doi:https://doi.org/10.1175/2007MWR2123.1, 2008.

Roche, D. M., Dokken, T. M., Goosse, H., Renssen, H., and Weber, S. L.: Climate of the Last Glacial Maximum: sensitivity studies and model-data comparison with the LOVECLIM coupled model, Clim. Past, 3, 205–224, doi:10.5194/cp-3-205-2007, 2007.

880

Scholes, D., and Hall, R.: The carbon budget of tropical savannas; woodlands and grasslands, in Global change: effects on; coniferous forests and grassland, edited by: Breymeyer, A. I., Hall, I.D., Melillo, J. M., and Agren, G. I., Wiley, New York, 69–100, 1996.

885 Schurgers, G., Mikolajewicz, U., Groeger, M., Maier-Reimer, E.,Vizcaıno, M., and Winguth, A.: Dynamics of the terrestrial biosphere, climate and atmospheric CO_2 concentration during interglacials: a comparison between Eemian and Holocene, Clim. Past, 2, 205–220, doi:10.5194/cp-2-205-2006, 2006.

Sitch, S., Smith, B., Prentice, I. C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J. O., Levis, S., Lucht, W., Sykes, M.
890 T., Thonicke, K., and Venevsky, S.: Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model, Glob. Chang. Biol., 9, 161–185, 2003.

Tang, G., Shafer, S.L., Bartlein, P., Holman, J.: Effects of experimental protocol on global vegetation model accuracy: a comparison of simulated and observed vegetation patterns for Asia. Ecol Model, 220(12), 1481–1491, 2009.

895

Taylor, K.E., R.J. Stouffer, G.A. Meehl: An Overview of CMIP5 and the experiment design. Bull. Amer. Meteor. Soc., 93, 485-498, doi:10.1175/BAMS-D-11-00094.1, 2012.

Tian , F., Cao, X., Dallmeyer, A., Lohmann, G., Zhang, X., Ni, J., Andreev, A., Anderson, P., Lozhkin, A., Bezrukova, E.,
900 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. Vegetation History and Archaeobotany, 27, 365-379. doi:10.1007/s00334-017-0653-8, 2018.

Torello-Raventos, M., Feldpausch, T. R., Veenendaal, E., Schrodt, F., Saiz, G., Domingues, T. F., Djagbletey, G., Ford, A.,
905 Kemp, J., Marimon, B. S., Hur Marimon Junior, B., Lenza, E., Ratter, J. A., Maracahipes, L., Sasaki, D., Sonké, B., Zapfack, L., Taedoumg, H., Villarroel, D., Schwarz, M., Quesada, C. A., Yoko Ishida, F., Nardoto, G. B., Affum-Baffoe, K., Arroyo, L., M. J. S. Bowman, D., Compaore, H., Davies, K., Diallo, A., Fyllas, N. M., Gilpin, M., Hien, F., Johnson, M., Killeen, T. J., Metcalfe, D., Miranda, H. S., Steininger, M., Thomson, J., Sykora, K., Mougin, E., Hiernaux, P., Bird, M. I., Grace, J.,

60

Lewis, S. L., Phillips, O. L., and Lloyd, J.: On the delineation of tropical vegetation types with an emphasis on forest/savanna transitions, Plant Ecol. Divers., 6, 101–137, https://doi.org/10.1080/17550874.2012.762812, 2013.

University of East Anglia Climatic Research Unit (Jones, P. D. and Harris, I. C.): Climatic Research Unit (CRU) time-series datasets of variations in climate with variations in other phenomena, NCAS British Atmospheric Data Centre, available at: http://catalogue.ceda.ac.uk/uuid/3f8944800cc48e1cbc29a5ee12d8542d (last access: November 2011), 2008.

University of East Anglia Climatic Research Unit; Harris, I.C.; and Jones, P.D.: CRU TS4.00: Climatic Research Unit (CRU) Time-Series (TS) version 4.00 of high-resolution gridded data of month-by-month variation in climate (Jan. 1901- Dec. 2015). Centre for Environmental Data Analysis, doi:10.5285/edf8febfdaad48abb2cbaf7d7e846a86. http://dx.doi.org/10.5285/edf8febfdaad48abb2cbaf7d7e846a86, 2017.

Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takemura, T., Okajima, H., Nozawa, T., Kawase, H., Abe, M., Yokohata, T., Ise, T., Sato, H., Kato, E., Takata, K., Emori, S., and Kawamiya, M.: MIROC-ESM 2010: model description and basic results of CMIP5-20c3m experiments, Geosci. Model Dev., 4, 845–872, https://doi.org/10.5194/gmd-4-845-2011, 2011

Wohlfahrt, J., Harrison, S.P., Braconnot, P., Hewitt, C.D., Kitoh, A., Mikolajewicz, U., Otto-Bliesner, B.L., and Weaver, S.L.: Clim Dyn, 31: 871, https://doi.org/10.1007/s00382-008-0415-5, 2008.

Wolff, J. K., Harrold, M., Fowler, T., Gotway, J.H., Nance, L., and Brown, B.G.: Beyond the basics: Evaluating model-based precipitation forecasts using traditional, spatial, and object-based methods. Wea. Forecasting, 29, 1451–1472, doi:https://doi.org/10.1175/WAF-D-13-00135.1, 2014.

Zhu, D.: Modeling terrestrial carbon cycle during the Last Glacial Maximum, Thèse de doctorat Météorologie, océanographie, physique de l'environnement Paris Saclay 2016, 2016SACLV077, http://www.theses.fr/2016SACLV077, 2016.

Zhu, D., Peng, S. S., Ciais, P., Viovy, N., Druel, A., Kageyama, M., Krinner, G., Peylin, P., Ottlé, C., Piao, S. L., Poulter, B., Schepaschenko, D., and Shvidenko, A.: Improving the dynamics of Northern Hemisphere high-latitude vegetation in the ORCHIDEE ecosystem model, Geosci. Model Dev., 8, 2263-2283, https://doi.org/10.5194/gmd-8-2263-2015, 2015.

.

PFT class	JSBACH	LPJ	ORCHIDEE	DGVM/CLM	TRIFFID	VECODE
Tropical trees	Tropical broadl. evergreen	Tropical broadl. evergreen	Tropical broadl. evergreen	Tropical broadl. evergreen	Broad-leaved trees & Needle-leaved trees	trees
	Tropical broadl. raingreen	Tropical broadl. raingreen	Tropical broadl. raingreen	Tropical broadl. raingreen		
Extra-tropical trees	Extra-tropical evergreen	Temperate needlel. evergreen	Temperate needlel. evergreen	Temperate needlel. evergreen		
		temperate broadl. evergreen	Temperate broadl. evergreen	Temperate broadl. evergreen		
		Boreal needlel. evergreen	Boreal needlel. evergreen	Boreal needlel. evergreen		
	Extra-tropical deciduous	Temperate broadl. deciduous	Temperate broadl. deciduous	Temperate broadl. deciduous		
		Boreal needlel. deciduous	Boreal needlel. deciduous	Boreal deciduous		
		Boreal broadl deciduous	Boreal broadl. deciduous			
shrubs	Raingreen shrubs				shrubs	
	Cold deciduous shrubs					
Grass	C3 grass	C3 grass	C3 grass	Artic grass / C3 grass	C3 grass	herbaceous
	C4 grass	C4 grass	C4 grass	C4 grass	C4 grass	

945   Tab.1: The plant functional types (PFT) used in the different state-of-the-art dynamic global vegetation models

Mega-biome	minimum coverage needed	Bioclimatic limitations
Tropical forest	Tropical trees dominant	$Tc > 15.5°C$
warm-temperate forest	Temperate trees dominant	$GDD5 > 3000°C$ *
Temperate forest	Temperate woody PFTs dominant	$900°C < GDD5 \leq 3000°C$ *
Boreal forest	Boreal woody PFTs dominant	$GDD5 \leq 900°C$ *
(Warm) savanna and dry woodland	Woody PFT coverage > 0.25	$GDD5 > 1200°C$ *, $Tc > 10°C$
Grassland and dry shrubland	Total vegetation cover > 0.2	$GDD0 \geq 800°C$ *
Tundra	Total vegetation cover > 0.1	$GDD0 < 800°C$ *
(Warm) desert	Total Vegetation cover < 0.2	$T_{ann} > 2°C$
Polar desert / ice	Total Vegetation cover < 0.2	$T_{ann} < 2°C$

Tab.2 Bioclimatic limits and assumptions on minimum PFT coverage needed for the assignment of PFTs into the PFT-groups and into the 9 Mega-Biomes. The separation of boreal, temperate and tropical tree-PFTs is based on the same bioclimatic limits as the respective forest mega-biomes. Only temperature-based limitations are used, i.e. the growing degree days on a basis of 5°C (GDD5) or on a basis of 0°C (GDD0), the monthly mean temperature of the coldest month (Tc), and the annual mean temperature (T_{ann}). Bioclimatic limits are mainly taken from the BIOME4 model (Kaplan et al., 2003, marked with *). The limit for tropical forest is taken from BIOME1 (Prentice et al., 1992), but is also commonly used in DGVMs (e.g. JSBACH). The limit for the differentiation of deserts has been empirically determined in this study and is close to the value chosen by Handiani et al. (2013) and within the range of the Koeppen-Geiger climate classification for polar climate and the Holdridge alpine life zone classification. The Tc limit for warm Savannas is taken from JSBACH (C4-grass criteria) to exclude temperate savannas. The assumptions on minimum coverage have partly been taken, partly been empirically adapted from Handiani et al. (2013). A flowchart displaying the biomisation procedure is shown in the Appendix A for the VECODE model, the DGVM used in Climber2.

Model acronym	Model (DGVM)	Period	Resolution	PFT	Reference
MPI-ESM-T63	MPI-ESM-P (JSBACH)	PI *#	T63	8+4	cmip5.output1.MPI-M.MPI-ESM-P. piControl.mon.land.Lmon.r1i1p1.v20120315 cmip5.output1.MPI-M.MPI-ESM-P. midHolocene.mon.land.Lmon.r1i1p2.v20120713 cmip5.output1.MPI-M.MPI-ESM-P. lgm.mon.land.Lmon.r1i1p2.v20120713
		6k			
		LGM			
MPI-ESM-T31	MPI-ESM-P (JSBACH)	PI, LGM	T31	8	Klockmann et al., 2016 (piCTL, LGMref)
IPSL-ESM-T31	IPSL-CM5A-LR (ORCHIDEE)	PI *#	1.875° x 3.75°	10+2	pmip3.output.IPSL.IPSL-CM5ALR.piControl. monClim.land.Lclim.r1i1p1.v20140428|
IPSL-ESM-T63	CRUNCEP or IPSL-CM5A-LR (ORCHIDEE-MICT)	PI, LGM	2°x2°	10	Zhu (2016)
HadGEM2-ESM	HadGEM2-ES (TRIFFID)	PI *	1.875° x 1.25°	6+2	cmip5.output1.MOHC.HadGEM2-ES. piControl.mon.land.Lmon.r1i1p1.v20111007 mip5.output1.MOHC.HadGEM2-ES.midHolo cene.mon.land.Lmon.r1i1p1.v20120222
		6k			
CLIM-LPJ	CRU/Climber2 (LPJ)	PI, 6k	0.5°x0.5	9	similar as in Kleinen et al. 2010
MIROC-ESM	MIROC-ESM (SEIB)	PI*	T42	8+2	cmip5.output1.MIROC.MIROC-ESM.piControl. mon.land.Lmon.r1i1p1.v20120710 Watanabe et al., 2011 cmip5.output1.MIROC.MIROC-ESM.midHolo cene.mon.land.Lmon.r1i1p1.v20120710 mip5.output1.MIROC.MIROC-ESM.lgm. mon.land.Lmon.r1i1p1.v20120710
		6k			
		LGM			
CLIMBER	Climber2 (VECODE)	PI, LGM	10°x10°	2	Kleinen (personal communication)

960

Tab.3: Overview of the simulations used for testing the biomisation method. Listed are the model acronym, the model name, the name of the included dynamic global vegetation model (DGVM), the simulations used in this study, the spatial resolution used in the simulations, the number PFTs (natural + anthropogenic) and the simulation reference. Simulations marked with (*) include land-use. Simulations marked with (#) ran with prescribed vegetation.

	PFT-based			Climate-based		
	PI	**6k**	**6k-PI**	**PI**	**6k**	**6k-PI**
MPI-ESM-T63	17.72	21.45	3.73	15.85	19.59	3.74
CLIM-LPJ	15.85	17.72	1.87	17.72	21.45	3.73
HadGEM2-ESM	13.99	15.85	1.86	13.99	15.85	1.86
MIROC-ESM	16.7	20.41	3.71	16.7	20.41	3.71

Tab.4 Position of the desert margin [° latitude] in North Africa at PI and 6k and the differences in position of the desert margin between 6k and PI [° latitude], for the PFT-based biomisations and the climate-based biomisations. The desert margin is here defined as latitude at which the zonal mean desert biome fraction averaged over the region 15°W to 30°E exceeds 50%.

	PFT-based			Climate-based		
	PI	**6k**	**6k-PI**	**PI**	**6k**	**6k-PI**
MPI-ESM-T63	0.73	0.75	0.02	0.73	0.76	0.03
CLIM-LPJ	0.64	0.76	0.12	0.67	0.77	0.1
HadGEM2-ESM	0.83	0.89	0.06	0.89	0.93	0.04
MIROC-ESM	0.72	0.72	0.0	0.92	0.96	0.04

Tab.5: Mean forest biome fraction in Northern Eurasia (0-150°E, 60-80°N) in the PFT-based biomisations and the climate-based biomisations for the mid-Holocene (6k) and pre-industrial (PI) time-slice, and the difference between both (6k-PI).

980

Biomes in Ramankutty and Foley (1999)	Mega-biomes
1 Tropical Evergreen Forest/Woodland	Tropical forest
2 Tropical Deciduous Forest/Woodland	
3 Temperate Broadleaf Evergreen Forest/Woodland	warm-temperate forest
4 Temperate Needleleaf Evergreen Forest/Woodland	Temperate forest
5 Temperate Deciduous Forest/Woodland	
6 Boreal Evergreen Forest/Woodland	Boreal forest
7 Boreal Deciduous Forest/Woodland	
8 Evergreen/Deciduous Mixed Forest/Woodland	Temperate (GDD5 < 900) or boreal forest
9 Savanna	Savanna and dry woodland, partly temperate forest
10 Grassland/Steppe	Grassland and dry shrubland
11 Dense Shrubland	Savanna and dry woodland
12 Open Shrubland	Grassland and dry shrubland
13 Tundra	Tundra
14 Desert	(Warm) Desert
15 Polar Desert/Rock/Ice	Polar desert / ice

Tab. C1 Biome assignment of biome classes used in Ramankutty and Foley (1999) to the mega-biomes used in this study.

Model acronym	RF99	PFT-based method	Climate-based method
MPI-ESM-T63	56.85	57.72	27.53
MPI-ESM-T31	58.52	61.0	31.13
IPSL-ESM-T31	59.75	40.28	7.87
IPSL-ESM-T63	57.51	59.03	59.58
HadGEM2-ESM	58.03	64.95	42.72
CLIM-LPJ	57.41	61.72	57.16
MIROC-ESM	58.45	53.08	15.45
CLIMBER	75.45	59.73	59.73

Tab. D1: Modern (PI) fraction of the biome 'tropical forest' in South America, given in % of land area in the region 33W-81W, 57S-13N. Listed are the reference forest fraction based on RF99, which change due to the interpolation to the different model grids, the forest fraction based on the PFT-biomisation and based on the classical approach using BIOME1 forced with the simulated climate fields.

	PI		LGM	
	FPC-method	**PFT-method**	**FPC-method**	**PFT-method**
FSS (vs RF99)	0.1	0.13	-	-
κ (vs RF99)	0.59	0.63	-	-
κ (vs. records)	0.19	0.24	0.17	0.13
BNS (vs. records)	0.53	0.55	0.54	0.41

Tab.E1: Metrics quantifying the agreement of the biomisations for IPSL-ESM-T63 based on the FPC-method (Prentice et al. 2011) or the PFT-based method introduced in this study with the modern potential biome distribution according to Ramakutty and Foley (1999, RF99) or pollen-based biome reconstructions (BIOME6000 database) for the pre-industrial (PI) and the Last Glacial Maximum (LGM) time-slices. Listed are the relative fractional skill score (FSS), the kappa value (κ) and the best neighbour score (BNS).

995

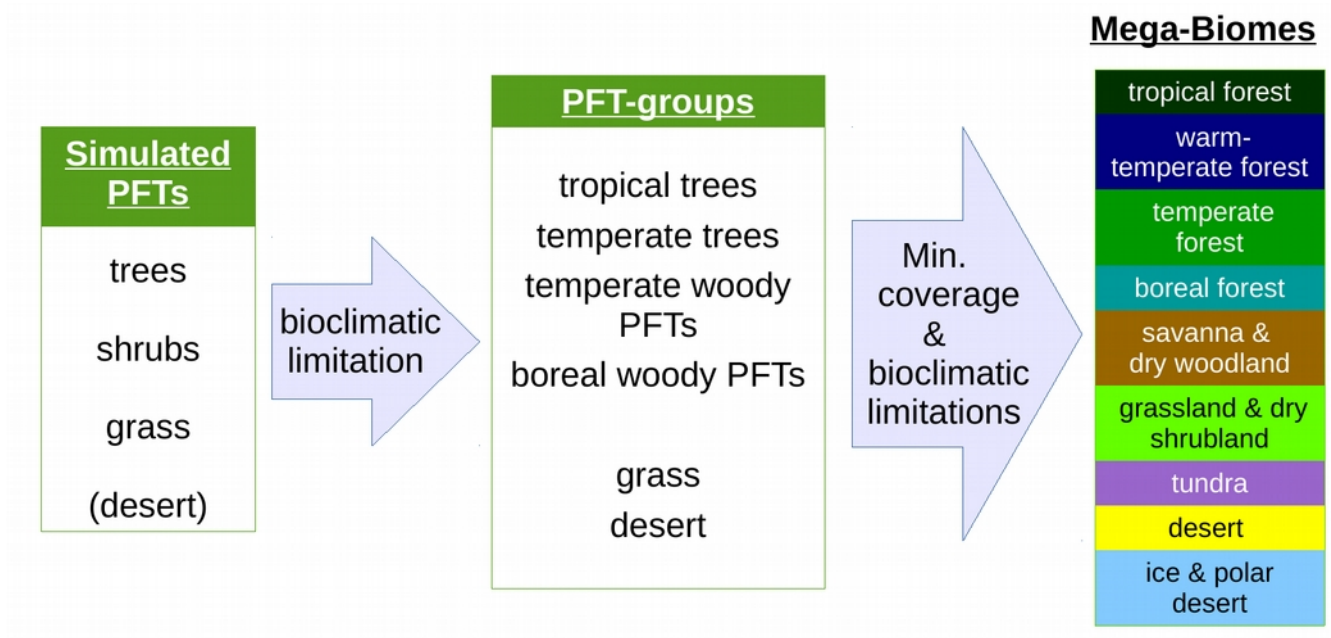



Figure 1: Scheme of the biomisation method. The plant functional type (PFT) fractions simulated by the individual dynamic global vegetation models (DGVM) are assigned to the PFT groups 'desert' (i.e. 1-total vegetation), 'grass' (containing all grass PFT-types) 'woody PFT' (containing all trees and shrub types) and 'trees' (containing all tree types). The 'trees' and 'woody PFTs' are further differentiated into 'tropical trees', 'temperate trees', 'temperate woody PFTs' and 'boreal woody PFTs' via bioclimatic limitations (Tab.2). For DGVM explicitly distinguishing tropical, temperate or boreal tree types, the original classification of the DGVM is used. Afterwards the PFT-groups are assigned to nine mega-biomes by assumptions on the minimum coverage of certain PFT-groups needed in a grid-cell and additional bioclimatic limitations (Tab.2).

1005

1010

80

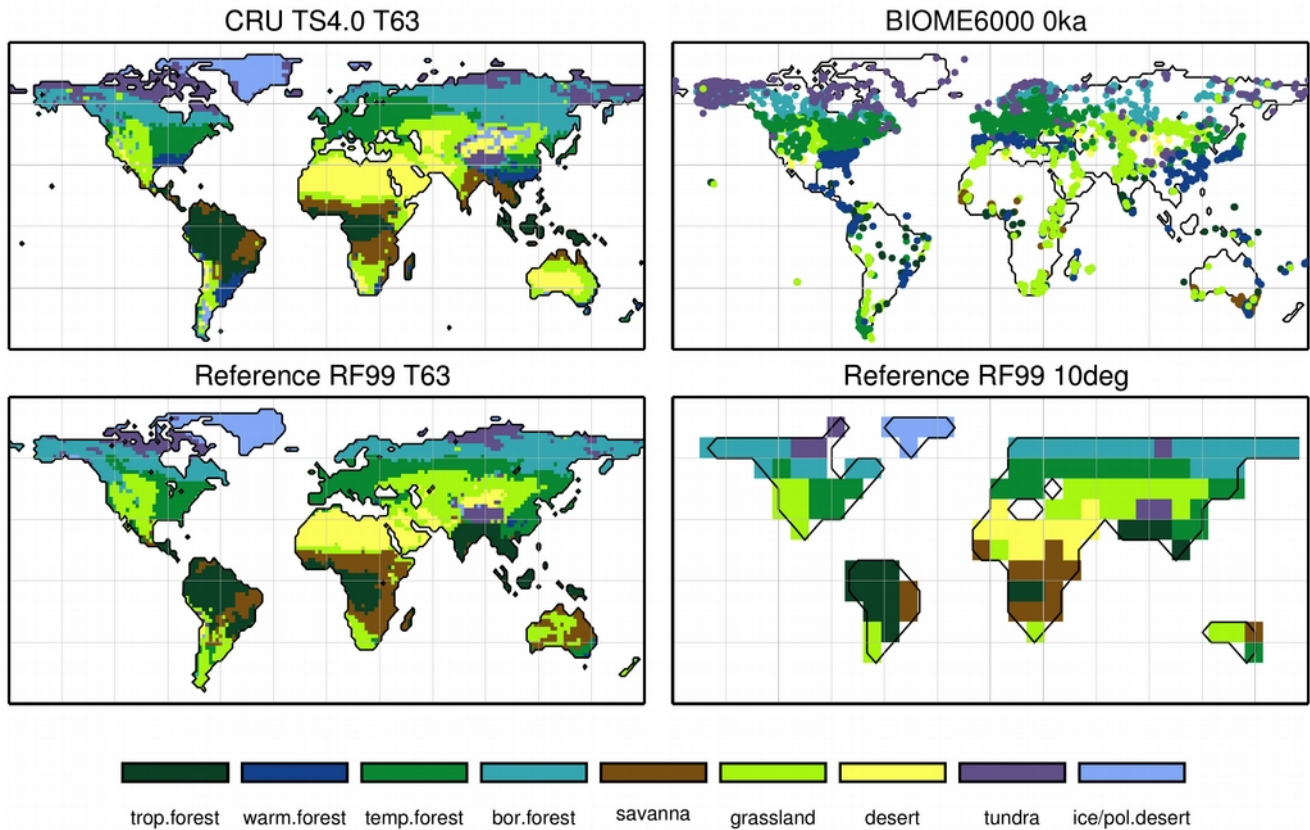



Figure 2: Reference biome distributions for the pre-industrial time-slice, i.e. upper left panel: the biome distribution inferred by BIOME1 that has been forced by the CRU TS 4.0 dataset (period 1901-1930), interpolated to a Gaussian T63 grid; upper right panel: the pollen-based pre-industrial biome reconstructions provided by the Biome6000 database (Harrison, 2017); lower panel: the modern potential natural vegetation map derived by Ramankutty and Foley (1999, RF99) remapped on a T63 gaussian grid (left) and a 10°grid (right). The biomes are tropical forest (trop.forest), warm-temperate forest (warm.forest), temperate forest (temp.forest), boreal forest (bor.forest), savanna & dry woodland (savanna), grassland & dry shrubland (grassland), warm desert (desert), tundra (tundra) and polar desert & Ice (ice/pol.desert).

1020

1025

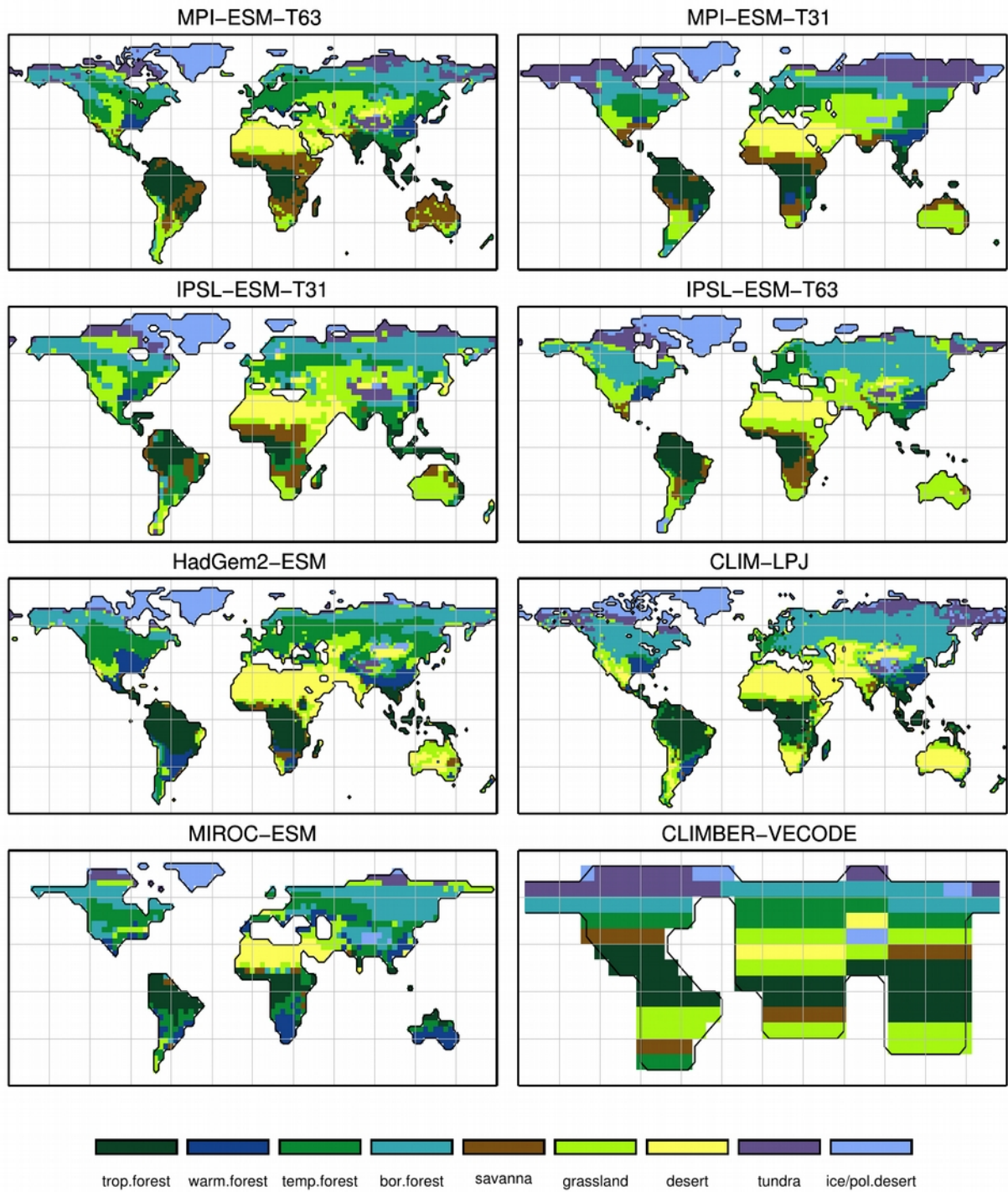



Figure 3: Simulated pre-industrial mega-biome distributions according to the new biomisation method (PFT-based method). The plant functional type (PFT) fractions simulated by the individual models have been converted into mega-biomes through climate limitation rules and assumptions on the maximum coverage of certain PFTs needed in the grid-cells.

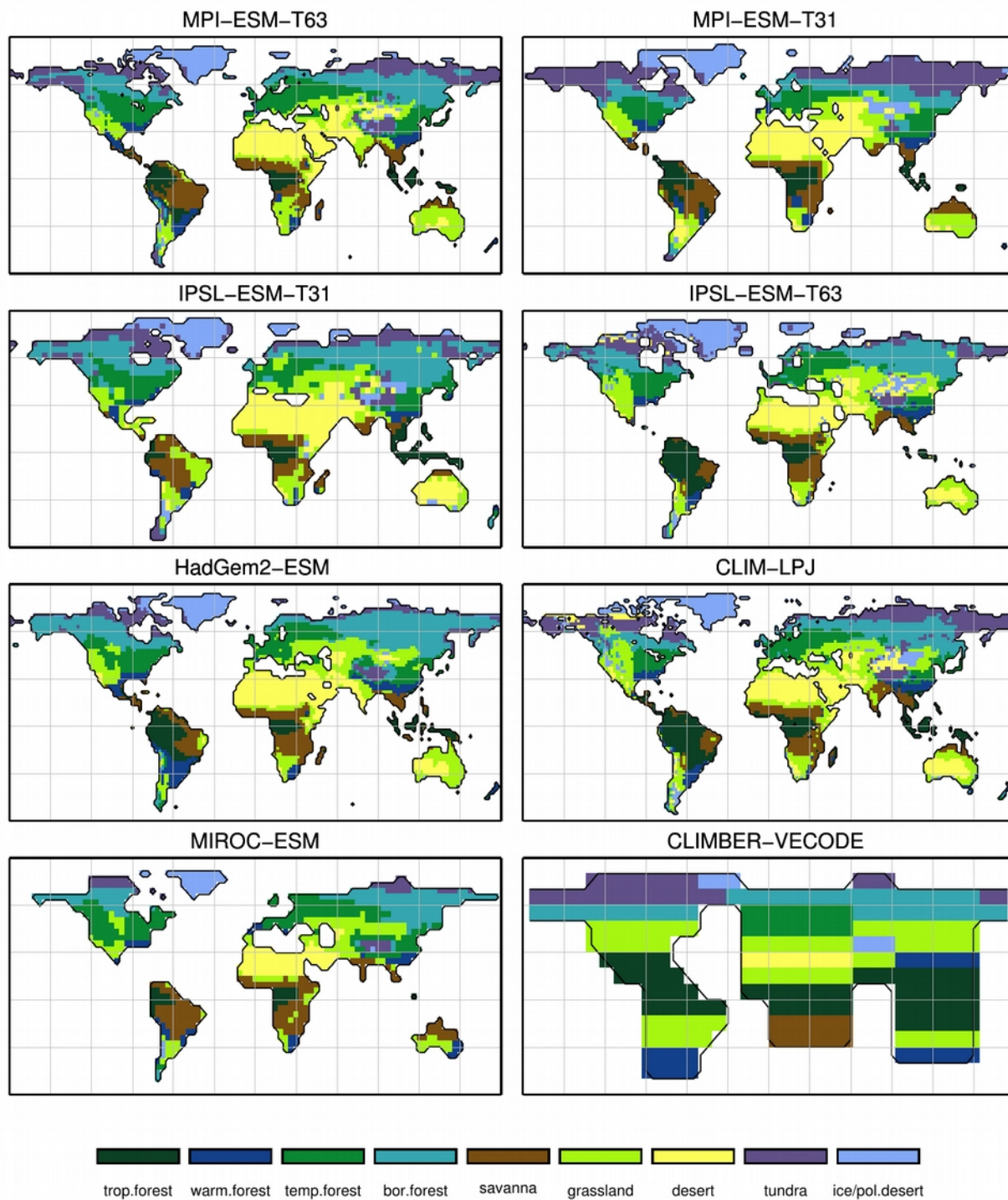



Figure 4: Simulated pre-industrial biome distributions according to the classical biomisation approach, i.e. biomising the climate states simulated by the individual models. The climate field were used to force the biome model BIOME1 (Prentice et al. 1992). Afterwards, the original BIOME1 biomes havebeen aggregated into the nine mega-biomes used in this study.

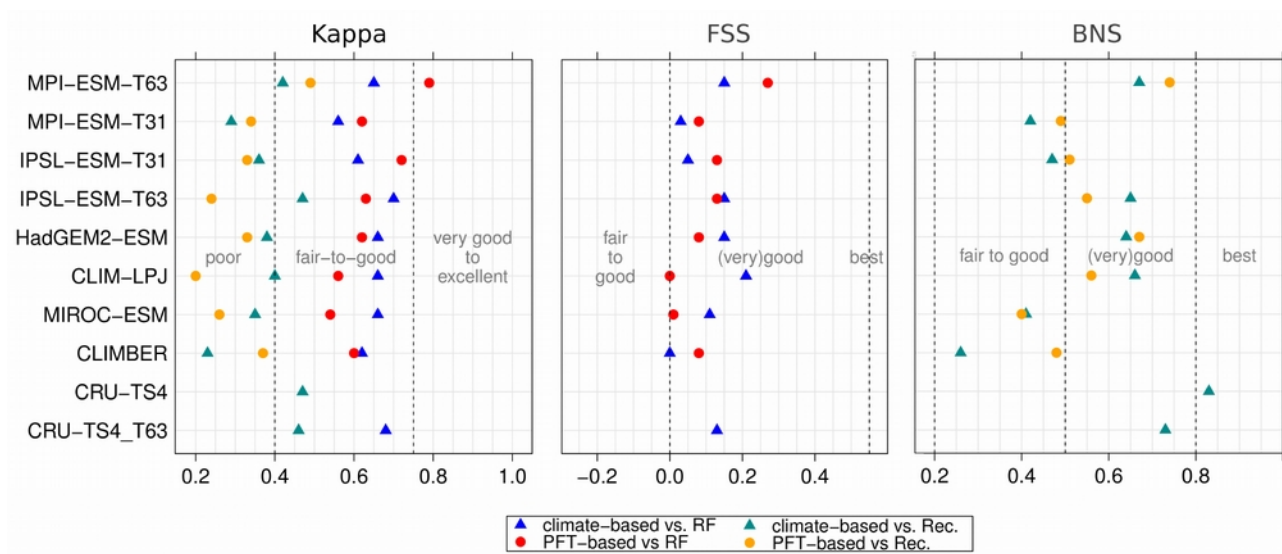



Figure 5: Metrics quantifying the total agreement of the simulated pre-industrial biome maps based on the PFT cover fractions (PFT-method) or based on the climate state (classical approach using BIOME1) with the reference datasets (i.e. modern potential natural vegetation (RF) and pre-industrial pollen-based biome reconstructions (Rec.)). Shown are the Kappa values (left figure), the relative fractional skill score (FSS, middle figure) and the best neighbour score (BNS, right figure) for all models and also for the biomisation based on the CRU-TS4 observational climate data in original resolution (CRU-TS4) and interpolated to a T63 grid (CRU-TS4_T63).

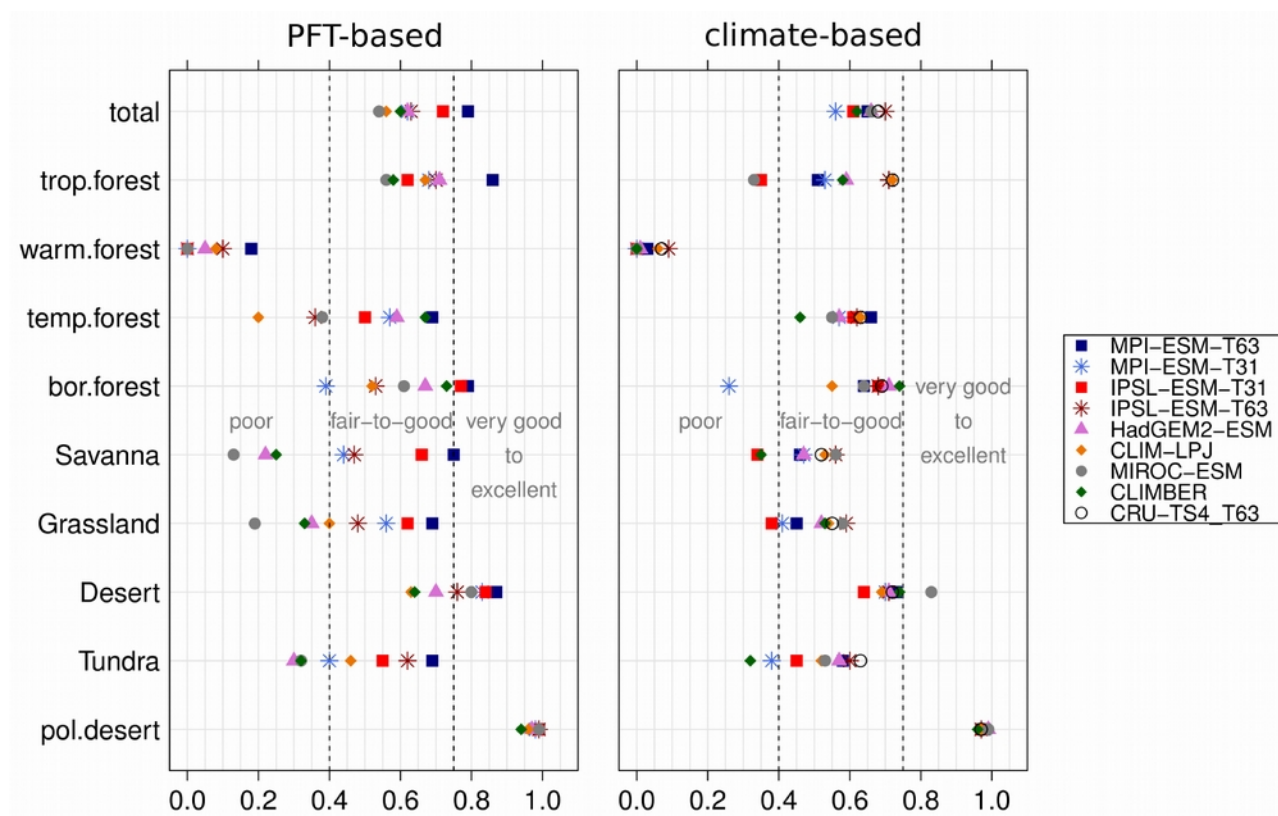



Figure 6: Kappa metric quantifying the agreement of the simulated pre-industrial individual mega-biomes with the reference dataset (i.e. modern potential natural vegetation (RF)) for the PFT-based method (left panel) and the classical method using BIOME1 forced with the simulated background climate states (right panel).

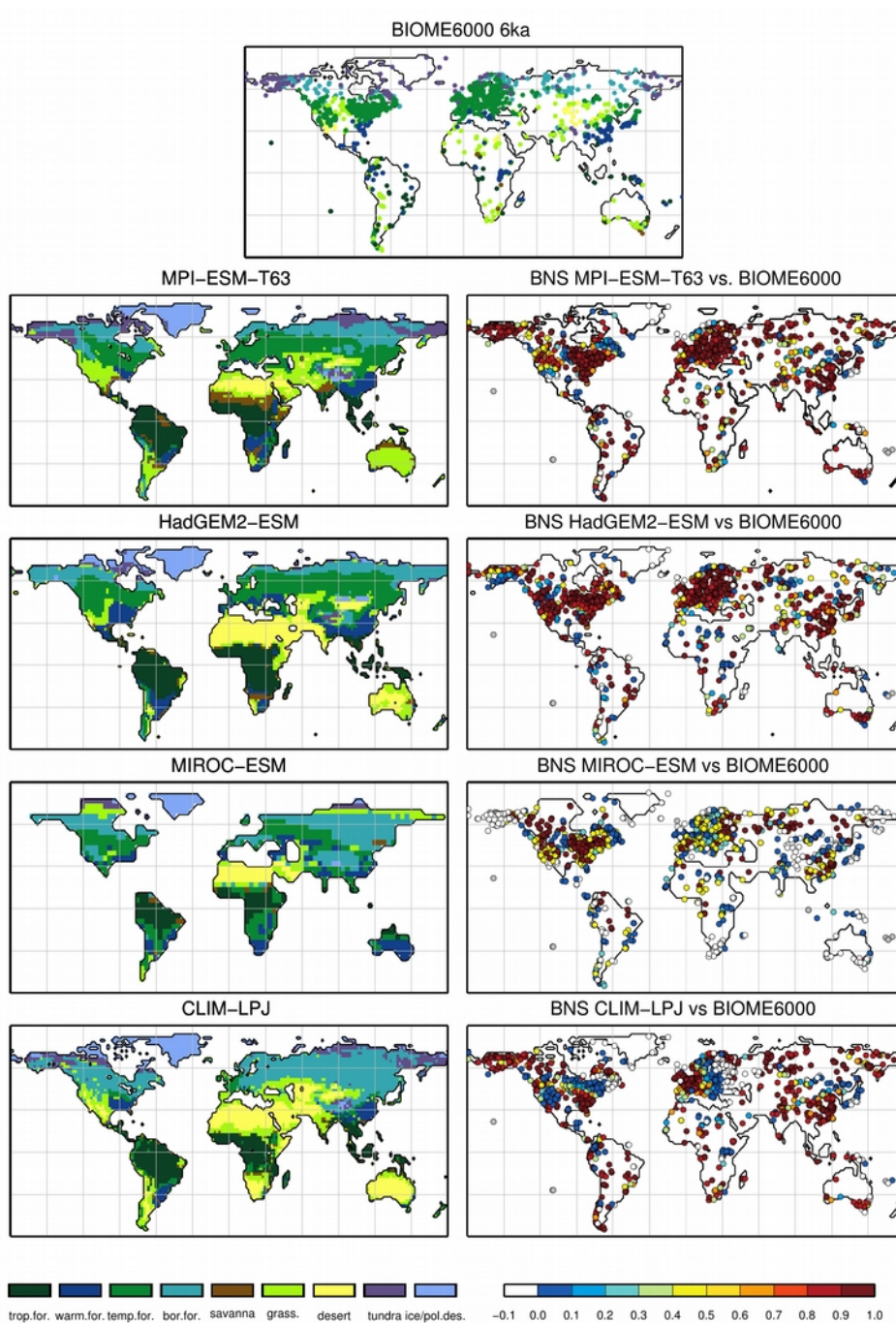



Figure 7: Simulated mid-Holocene biome distribution in the different models, based on the PFT-method (left panel), pollen-based biome reconstructions of the mid-Holocene biome distribution (BIOME6000 database, upper figure) and the best neighbour score (BNS) for all individual sites showing the agreement of the reconstructed biomes and the biome distribution in the neighbourhood of the sites, ranging from 0 (no grid-cell in the surrounding shows the same biome as reconstructed) to 1 (the grid-cell locating the site and the record at the site indicate the same biome).

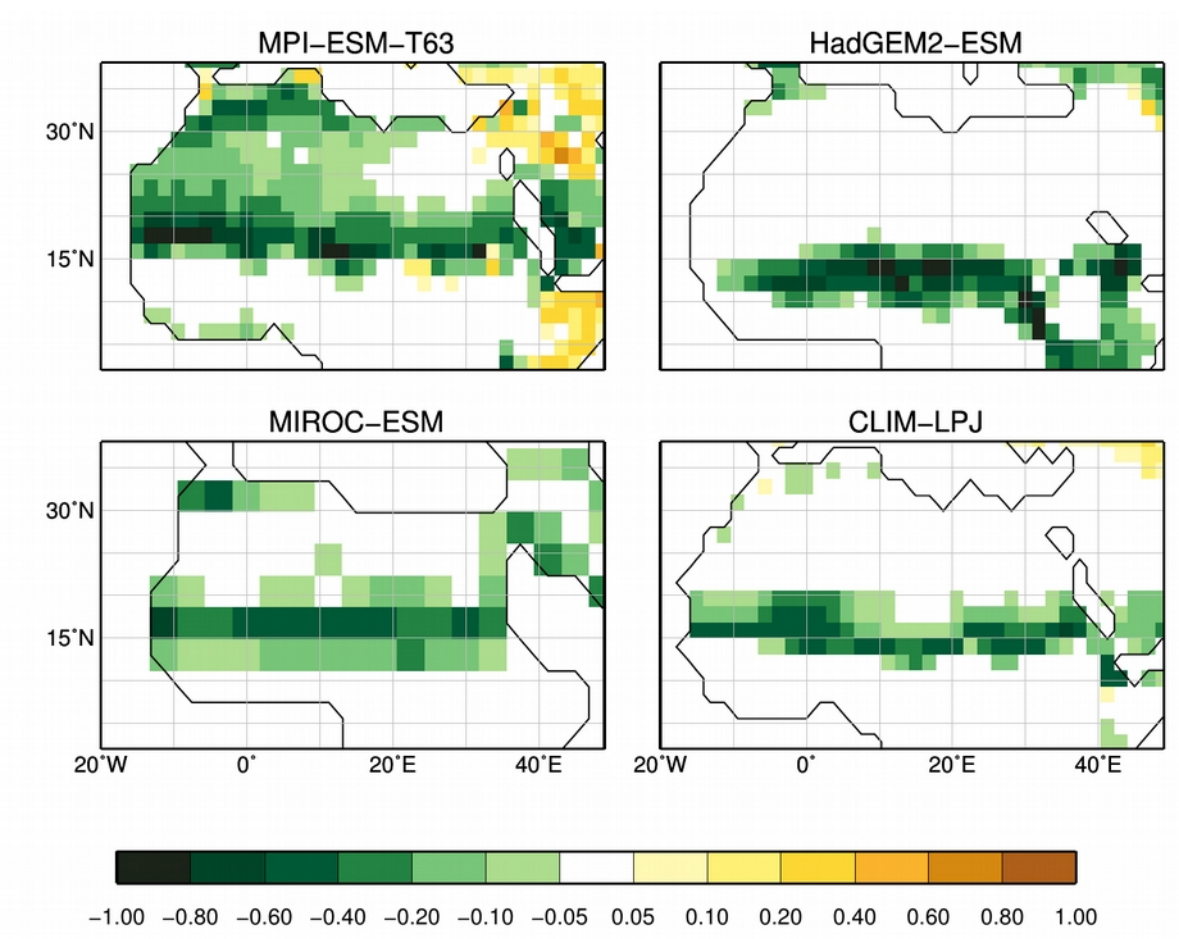



Figure 8: Differences in the desert fractional coverage simulated by the individual models between the mid-Holocene (6k) and pre-industrial time-slice (0k).

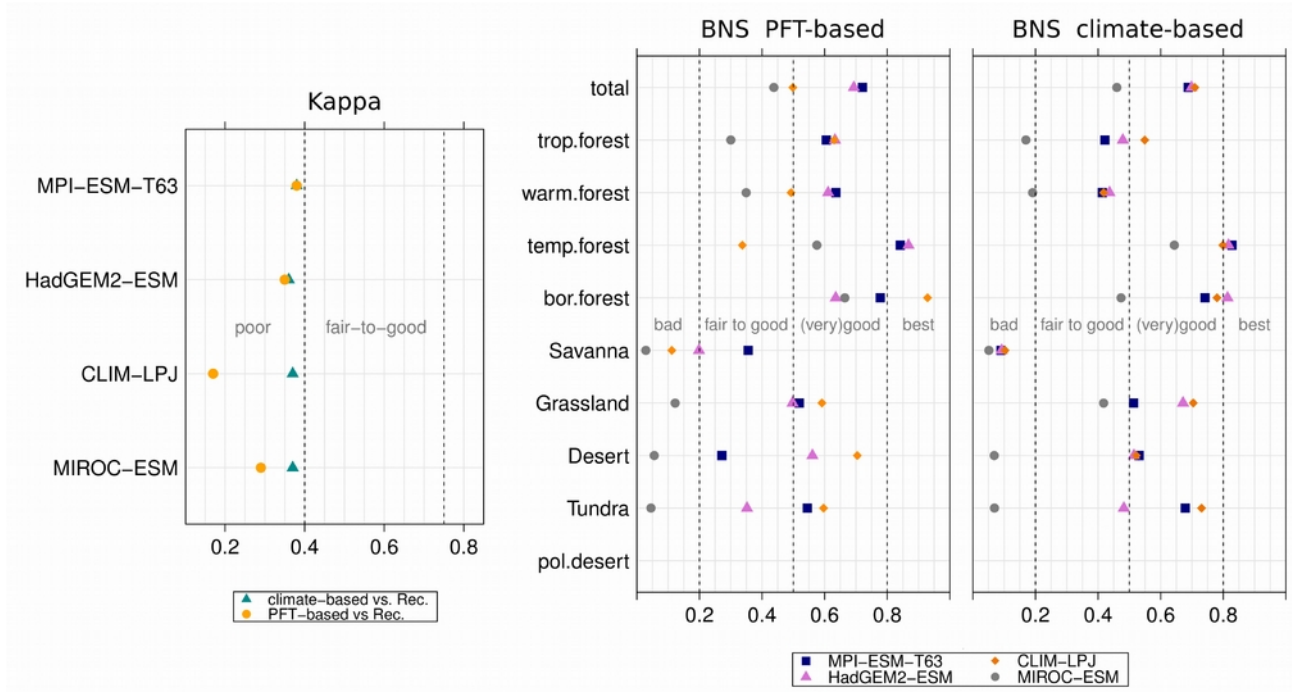



Figure 9: Metrics quantifying the agreement of the simulated mid-Holocene biome maps based on the PFT-method or based on the climate states (i.e. according to the BIOME1 model)with the pollen-based biome reconstructions (BIOME6000 database) for the mid-Holocene time-slice, i.e. the (total) Kappa value (left panel) and the BNS values for the individual mega-biomes.

1065

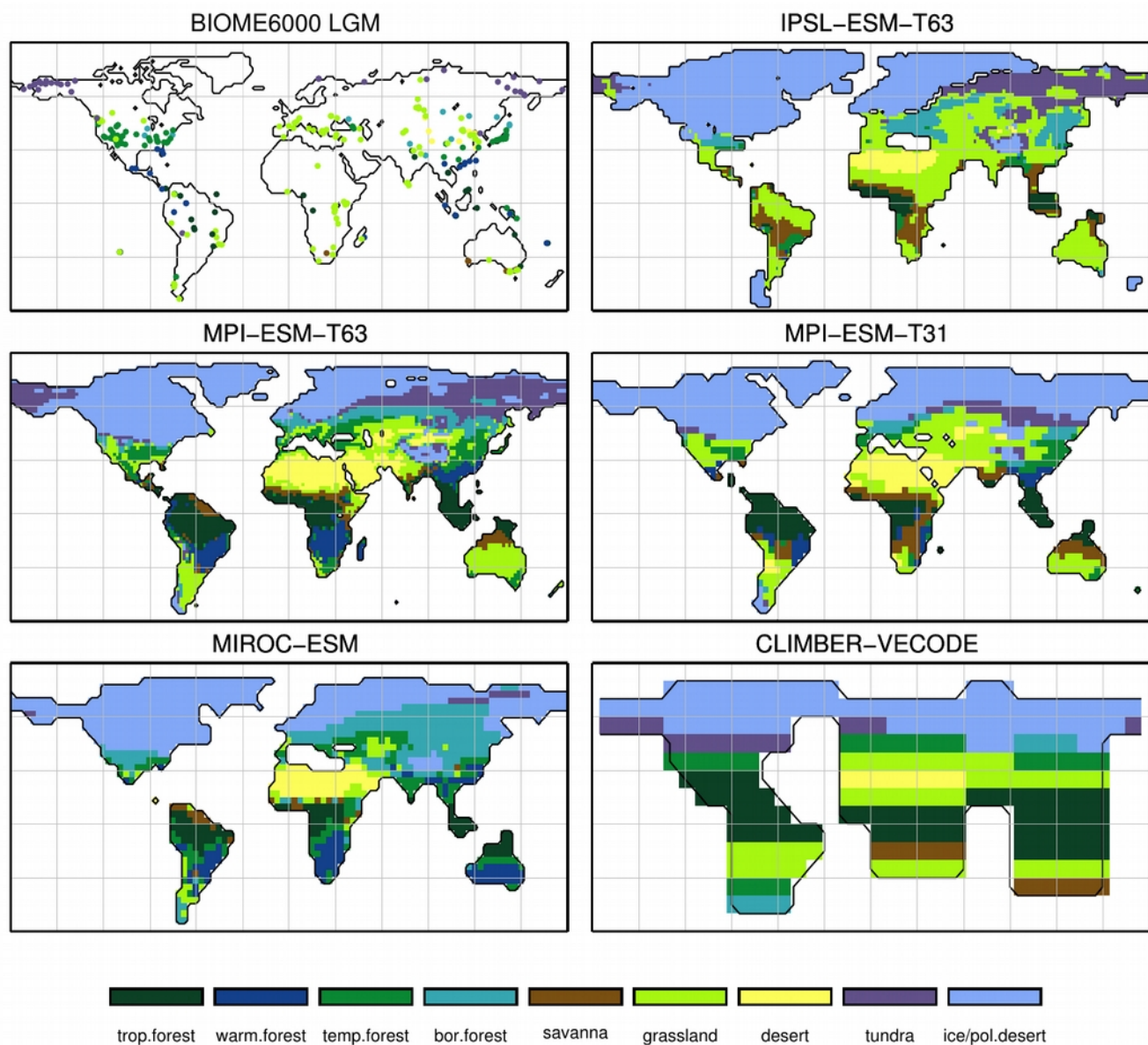



Figure 10: Pollen-based biome reconstructions (BIOME6000 database) for the Last Glacial Maximum time-slice and the simulated biome distributions according to the new biomisation method (i.e. the PFT-based method)

1070

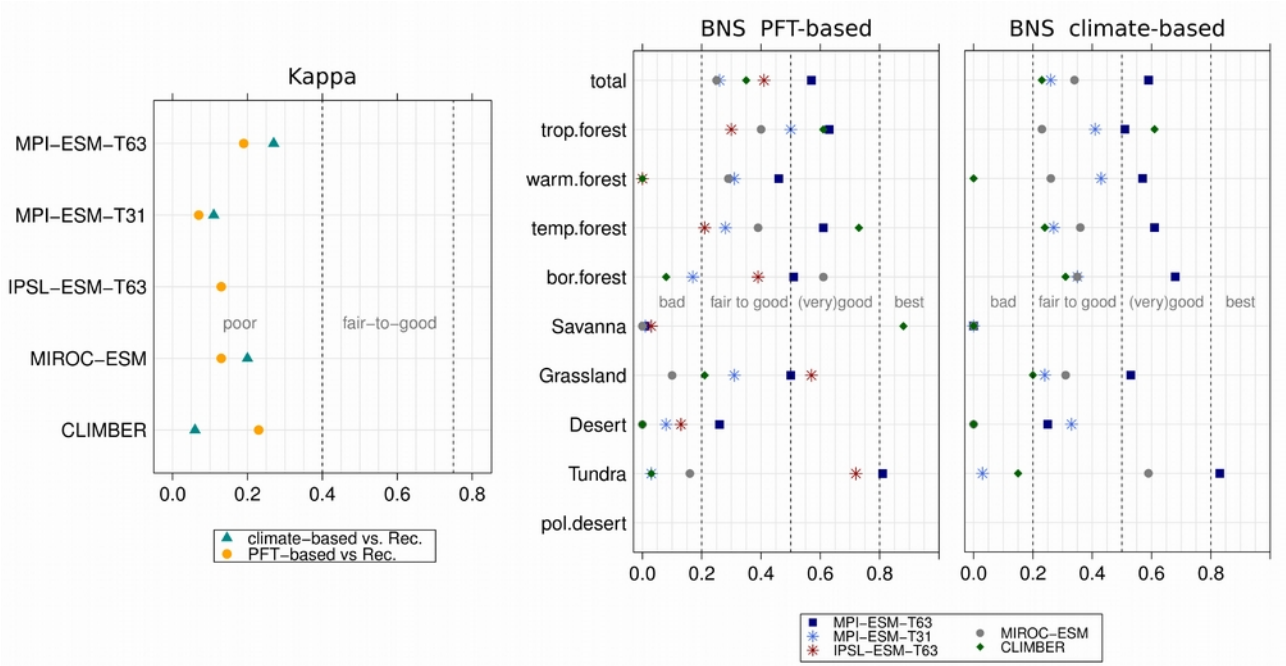
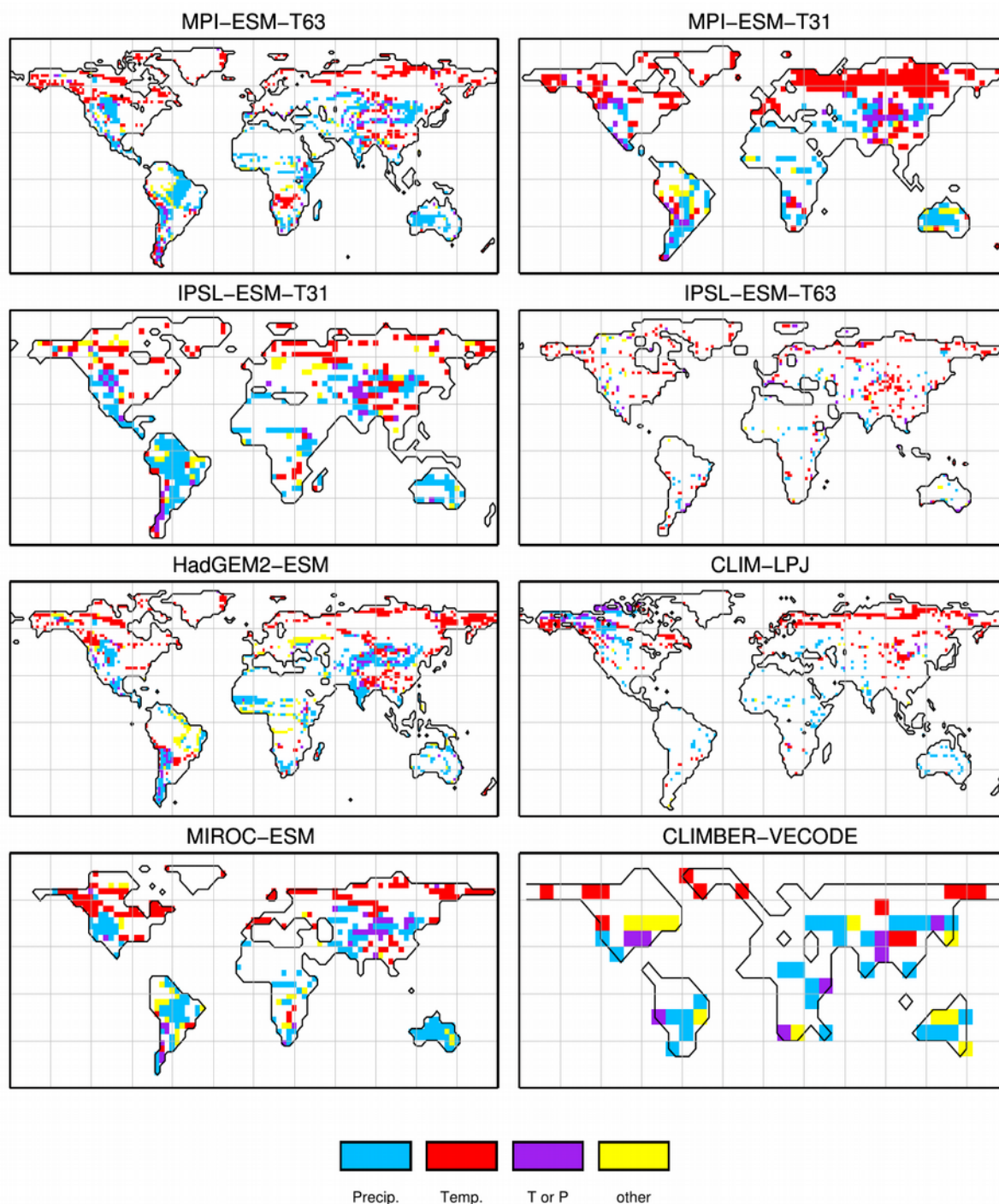



Figure 11: Metrics quantifying the agreement of the simulated Last Glacial Maximum biome maps based on the PFT-method or based on the simulated climate states (i.e. according to BIOME1) with pollen-based biome reconstructions (BIOME6000 database), i.e. the (total) Kappa value and the BNS values for the individual mega-biomes. Please notice that the climatic variables needed to force BIOME1 could not be provided for IPSL-ESM-T63. Thus, no climate-based biomisation exists for IPSL-ESM-T63.

1075

1080   **Figure 12: Climate factors leading to the differences between the pre-industrial climate-based biome distributions and the biome distribution inferred from the CRU-TS4.0 observational climate data. The factors were calculated by performing a sensitivity test with the BIOME1 model following Dallmeyer et al. (2017) by successively replacing the temperature or the precipitation in the CRU-TS4.0 forcing file for BIOME1 with the respective data from the different PI siumlations.**

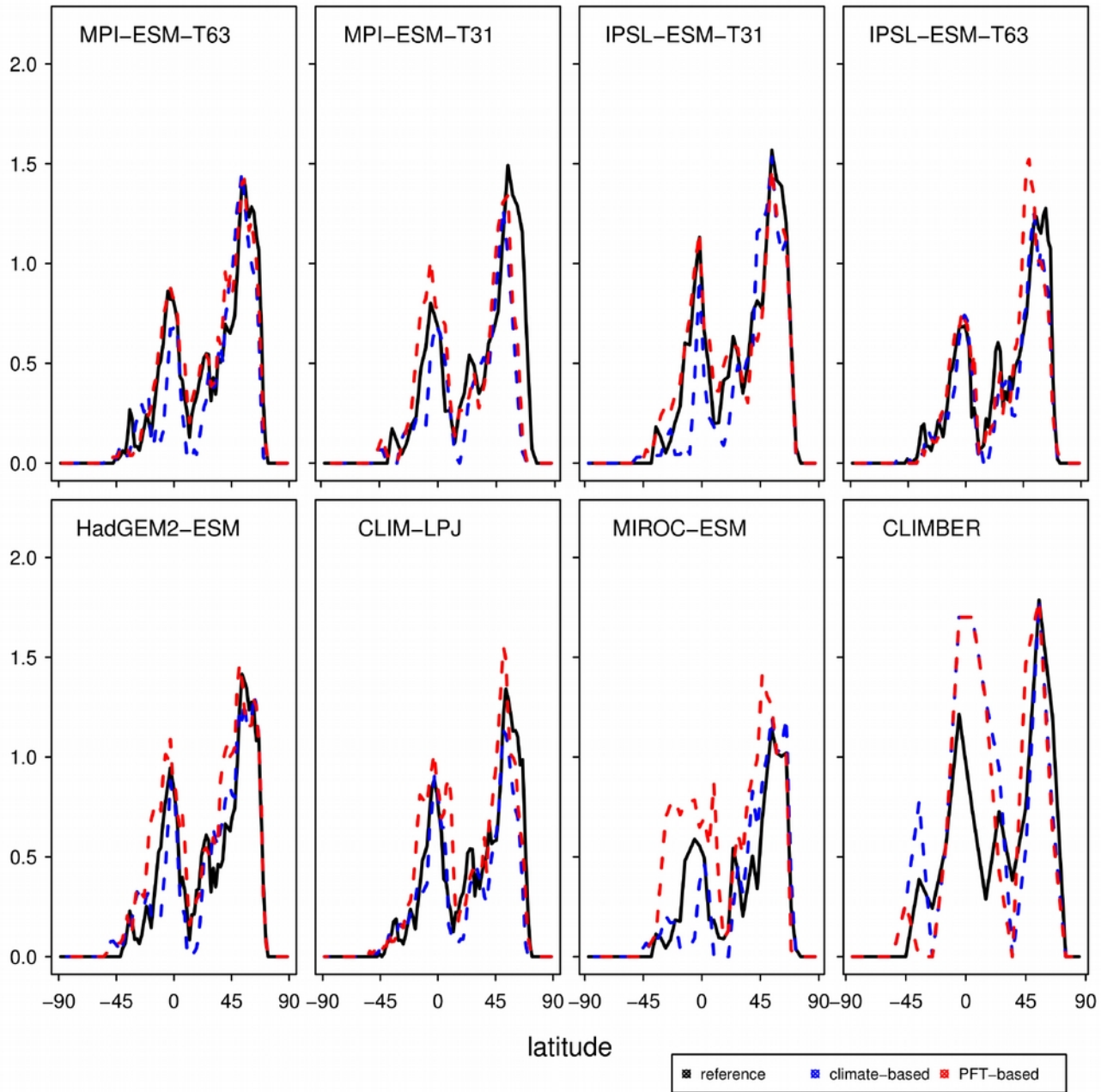



Figure 13: Zonal sum of pre-industrial forest biome area per latitude [Mio km²/degree] in the reference (RF99), the climate-based biomisation (blue) and the PFT-based biomisation (red) for each of the individual models. Due to the special land-sea mask in CLIMBER, the values for this model have been scaled by a factor of 0.766, which is the quotient of the global land area in a Gaussian T31 grid and the global land area in the CLIMBER grid (10°x10°).

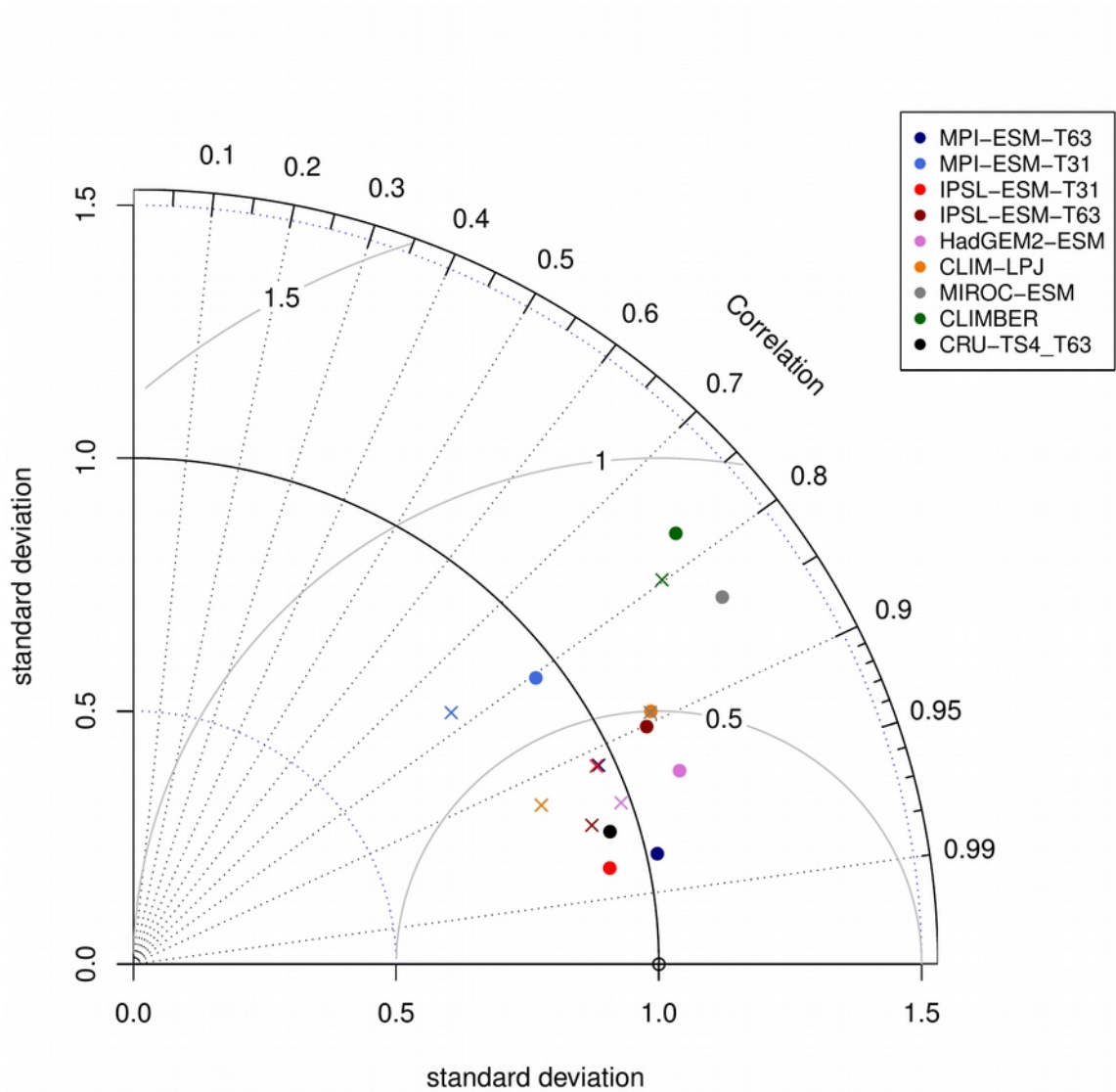



Figure 14: Normalised Taylor-diagramm showing the agreement of the simulated pre-industrial zonal sum of forest biome area per latitude using BIOME1 (i.e. based on the simulated climate states, crosses) or the PFT-based method (dots) for the individual models with the modern potential biome distributions according to Ramakutty and Foley (1999, RF99). Additionally shown is the agreement of the CRU-TS4-based biomisation with this RF99 reference dataset.

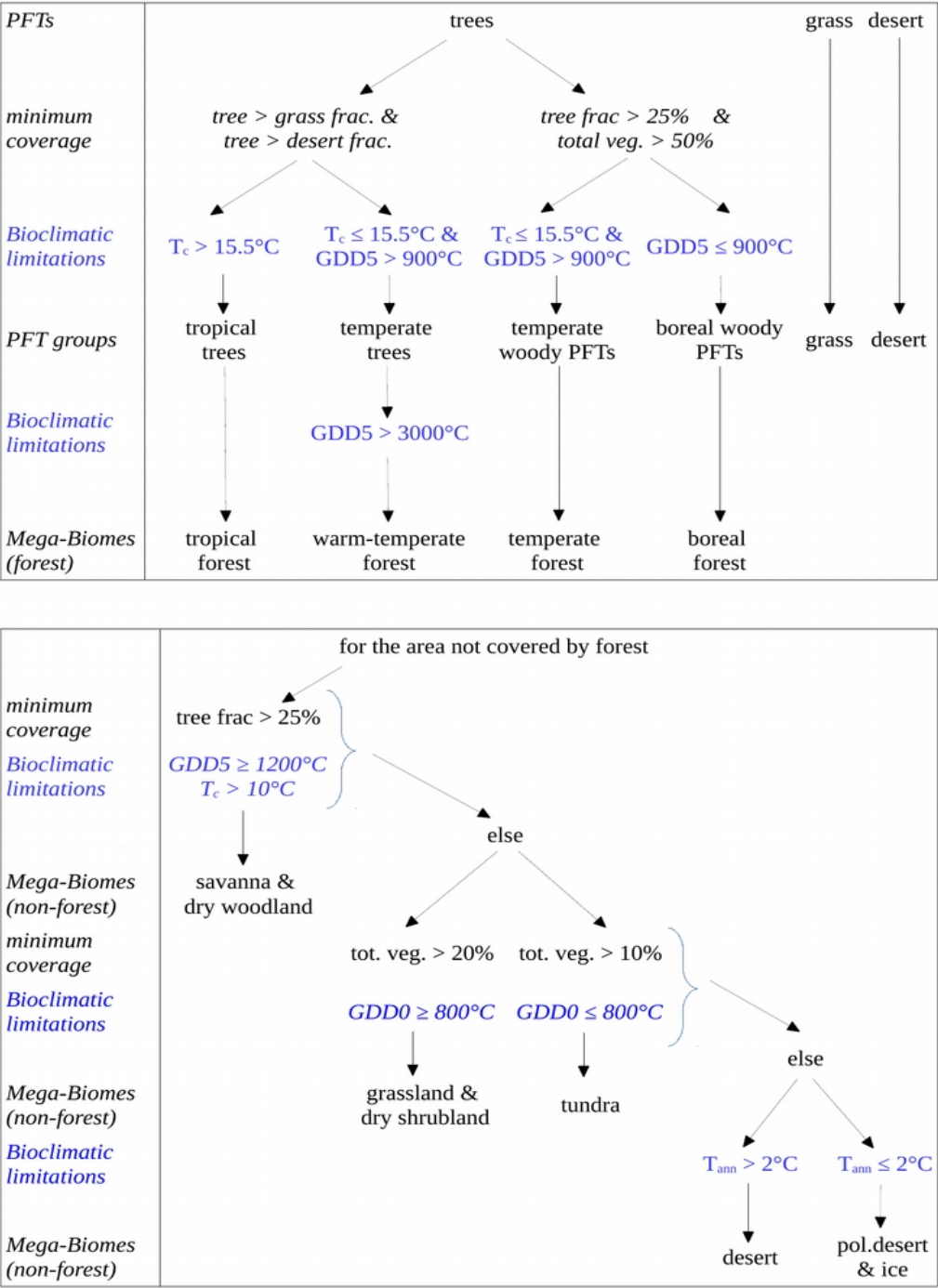



Figure A1: The PFT-biomisation procedure using the example of the VECODE model (included in Climber2), which has the fewest PFTs. Shown are all decisions using assumptions on minimum coverage of the PFTs and bioclimatic limitations, i.e. growing degree days on the basis of 0° and 5° (GDD0, GDD5), mean temperature of the coldest month (Tc) and annual mean 2m- temperature. Please notice, that this chart looks different for every other model.

1100

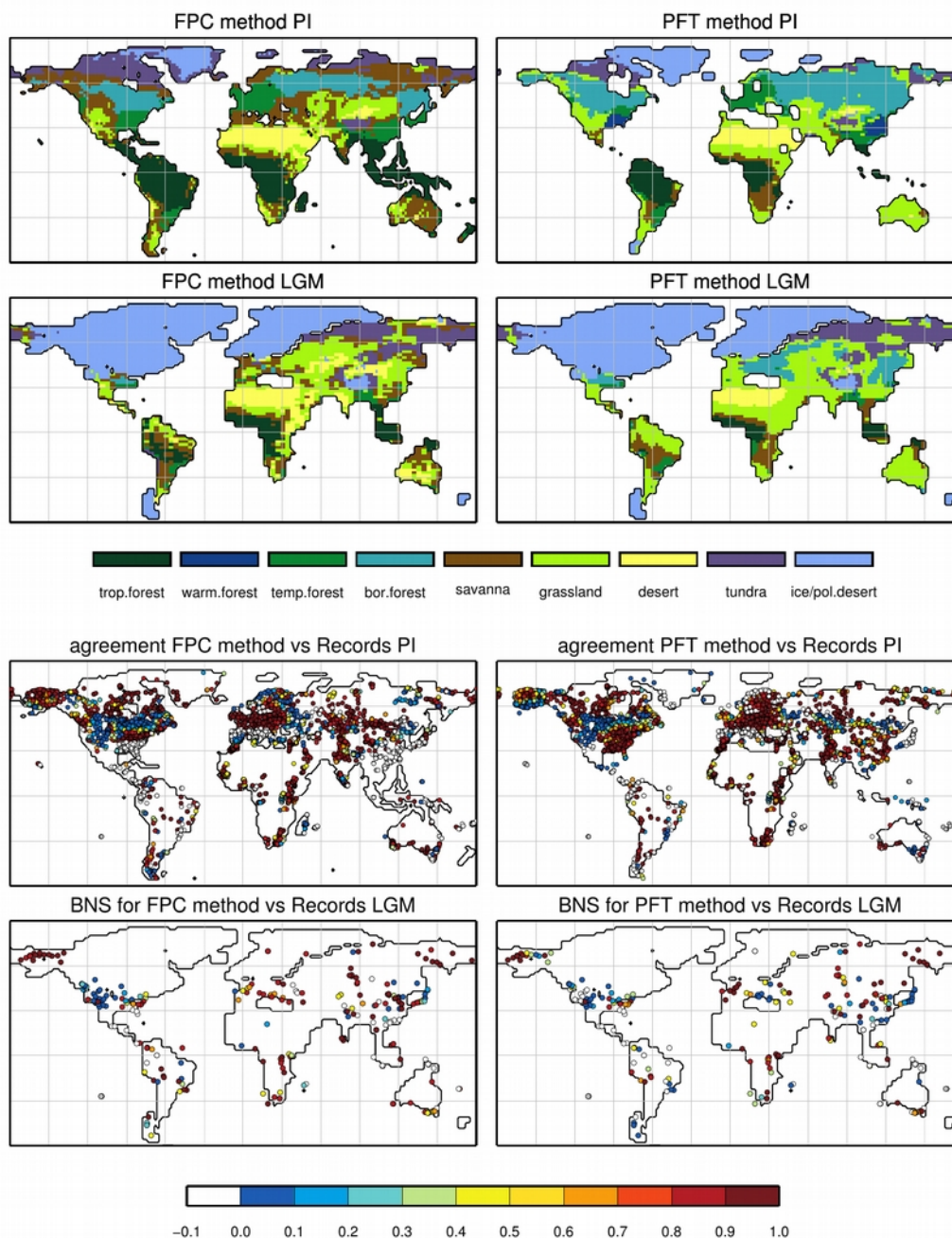



Figure E1: Comparison between the biomisation method by Prentice et al. (2011, referred to as FPC-method) and the method introduced in this study (PFT-based) based on the model IPSL-ESM-T63. Shown are the biome distributions for the pre-industrial (PI) and Last Glacial Maximum (LGM) time-slice (upper panels) and the BNS for all available sites showing the agreement of the reconstructed biomes (according to the BIOME6000 database) and the simulated biome distributions in the neighbourhood of the sites (lower panels), ranging from 0 (no grid-cell in the surrounding shows the same biome as reconstructed) to 1 (the grid-cell locating the site and the record at the site indicate the same biome). Please notice that the FPC-method distinguishes temperate parkland, sclerophyll woodland and boreal parkland that all have been assigned in the mega-biome savanna, but savanna is only defined as tropical savanna in the PFT-method.

1105

110