Crowd-sourced trait data can be used to delimit global biomes

Simon Scheiter[1], Sophie Wolf[2], and Teja Kattenborn[3]

[1]Senckenberg Biodiversity and Climate Research Centre (SBiK-F), Senckenberganlage 25, 60325 Frankfurt am Main, Germany
[2]Remote Sensing Centre for Earth System Research (RSC4Earth), Leipzig University and Helmholtz Centre for Environmental Research, Talstr. 35, 04103 Leipzig, Germany
[3]Department for Sensor-based Geoinformatics, Faculty of Environment and Natural Resources, University of Freiburg, Tennenbacherstr. 4, 79106 Freiburg, Germany

Correspondence: Simon Scheiter (simon.scheiter@senckenberg.de)

Abstract. Biomes and their biogeographic patterns have been derived from a large variety of variables including species distributions, bioclimate or remote sensing products. Yet, whether plant trait data are suitable for biome classification has rarely been tested. Here, we aimed to assess systematically which traits are most suitable for biome classification. We derived patterns of 33 different traits by combining crowd-sourced species distribution data and trait data from the TRY database.

5   Using supervised cluster analyses, we developed biome classification schemes using these traits and 31 different biome maps. A sensitivity analysis with randomly sampled combinations of traits was performed to identify traits and biome maps that are most appropriate for biome classification and achieved the highest data-model agreement. Due to gaps in the trait data, species distribution models were used to obtain biome maps at the global scale. We showed that traits can be used for biome classification and that the most appropriate traits are conduit density, rooting depth, height, and different leaf traits, including

10 specific leaf area and leaf nitrogen. Data-model agreement was maximized when biome maps used to inform cluster analyses were based on biogeographic zonation and species distributions, in contrast to biome maps derived from optical reflectance. The availability of crowd-sourced trait data is heterogeneous and large data gaps are prevalent. Nonetheless, it is possible to derive biome classification schemes from these data to predict global biome patterns with good agreement. Filling data gaps is essential to further improve trait-based biome maps.

1   Introduction

15 Biomes are commonly used to represent major vegetation formations and to map their biogeographic distributions. Multiple biome maps were developed based on a variety of different data sources Beierkuhnlein and Fischer (2021). These include biogeographic zonation based on species distributions, bioclimatic and edaphic variables, or a variety of remote sensing products such as NDVI, reflectance or vegetation height. Recently, Fischer et al. (2022) aggregated 31 different biome and land cover

20 maps, showing considerable differences between classification methods, the biome types represented in the maps and their spatial distributions. In addition to species distribution data or remote sensing data, plant traits provide a detailed representation of vegetation at the site level. It has been argued, that traits should be included in biome classification (Mucina, 2019; Hunter et al., 2021), and previous studies showed that trait data are suitable for delineating biomes (van Bodegom et al., 2014;

Boonman et al., 2022; Scheiter et al., 2024). Despite the increasing availability of trait data in databases such TRY (Kattge
25   et al., 2020) and extrapolated global biome maps Wolf et al. (2022); Boonman et al. (2020), a systematic assessment of the
performance of traits for biome classification and an identification of the most appropriate traits remain elusive.

In a recent study, Boonman et al. (2022) used height, specific leaf area (SLA) and wood density to reproduce the Olson et al.
(2001) biome map by applying a supervised cluster analysis. Global maps for these traits were derived by extrapolating trait
data from TRY using statistical and machine-learning approaches (Boonman et al., 2020). Scheiter et al. (2024) used traits
30   simulated by a dynamic vegetation model, the aDGVM2 (Langan et al., 2017; Kumar et al., 2021), and 31 different biome
maps provided by Fischer et al. (2022, hereafter F31 maps) for biome classification. While both studies showed that traits can
be used for biome classification, both approaches show caveats.

Boonman et al. (2022) used only a set of three different traits - height, wood density and SLA - even though data on more
traits are available in TRY (Kattge et al., 2020). The selection of traits was mainly driven by methodological criteria rather than
35   by ecological knowledge, and ensured that traits could be extrapolate to the global scale with sufficient predictive accuracy
using their statistical method. Extrapolation of geographically sparse TRY data to the global scale is uncertain (Ludwig et al.,
2023), and Dechant et al. (2023) showed large differences between global trait maps created using different extrapolation
methods. Such uncertainties in trait maps may propagate to uncertainties in biome classification.

Trait maps simulated by the aDGVM2 are provided as continuous maps for the simulated study region (Scheiter et al., 2024).
40   Limitations of using model results are model uncertainties in the aDGVM2, mismatches between observed and simulated trait
patterns and, accordingly, uncertainties in biome classification. Further, some of the traits simulated by the aDGVM2 are
difficult to observe in reality, while other traits are not represented in models. Therefore, classification schemes derived from
aDGVM2 cannot be directly applied to real-world situations. These limitations may bias the suitability of different traits
modeled in Scheiter et al. (2024) for biome classification.

45   A further caveat of both studies is that traits of different plant functional types (PFTs) were averaged and not considered
separately in the classification. Given that traits of grasses and trees, for example, differ substantially, we expect that such
averaging reduces the performance of biome classification compared to a classification where traits of different PFTs are
considered separately.

Recently, Wolf et al. (2022) presented a novel approach to create large-scale maps for a series of traits by merging crowd-
50   sourced species distribution data from iNaturalist with trait data from the TRY database (Kattge et al., 2020). To enhance the
spatial extent of trait data, this approach was now extended by integrating species distribution data from the Global Biodiversity
Information Facility (GBIF). Moreover, trait patterns were derived separately for grasses and woody plants, and for all PFTs
combined. We used these trait maps and supervised cluster analyses to create trait-based biome maps informed by all F31
observation-based biome maps compiled by Fischer et al. (2022). We asked: (1) are trait patterns derived by combining crowd-
55   sourced species distribution data and the TRY trait data appropriate to delimit global biomes despite substantial data gaps? (2)
Which and how many traits are appropriate and required to delimit global biomes? (3) Does the consideration of PFT-specific
traits improve biome classification? (4) Which trait values are characteristic for different biomes across all F31 biome maps?

The trait data were not available for the entire global land surface. To obtain global trait-based biome maps, we applied species distribution models for extrapolation.

2 Materials and Methods

2.1 Data

We used trait patterns derived from combining GBIF species observation data and trait data from the TRY database (Kattge et al., 2020, vers. 6). The methodology follows Wolf et al. (2022), but instead of using iNaturalist data we used the entire GBIF archive to increase geographic and taxonomic coverage. Species observations ($n = 257,357,303$) were downloaded from GBIF on June 2nd, 2023 (GBIF.Org User, 2023). This download already included filtering for 'Observation', 'Human observation' or 'Occurrence' records with no geospatial issues, no related records, a minimum distance of 1500 m to a country centroid and the occurrence status set to 'present', since GBIF contains true absence data. Additional filtering using the package R 'CoordinateCleaner' (Zizka et al., 2019) removed records with a coordinate uncertainty > 10km and a precision $> 0.1°$, those located in the ocean and those that match common issues, such as records that are falsely given coordinates along the equator or central meridian, and made sure to only include records identified to the species level. The observations were then linked to the TRY gap-filled dataset via species names (Schrodt et al., 2015), which resulted in a total of $n = 192,667,225$ observations. 90% of the filtered GBIF observations and 24% of species in GBIF were matched with 70% of species in TRY (numbers based on map products using all plant functional types). We spatially subsampled the data to limit both computational load and data redundancy: the matched observations were binned into equal area hexagons of 2,591 km^2 using the package 'dggridR' (Barnes and Sahr, 2023), which corresponds to about 0.5° at the equator. From each hexagon, we then sampled 10,000 observations. If a hexagon contained less than 10,000 observations, all observations were kept. After this subsampling 31,808,221 observations remained. These trait observations were then aggregated using a mean function to a global raster grid with 0.5°spatial resolution. We created three different maps for each trait in this manner, each for different combinations of plant functional types (PFTs): 1) grasses and herbs (hereafter grasses), 2) trees and shrubs (hereafter woody plants) and 3) all PFTs (i.e., grasses and woody plants). Therefore, we filtered the observations according to PFTs before spatially aggregating the trait values, where the association of a plant species to PFT was based on a majority vote based on PFT information contained in the TRY database (trait ID 197). All traits included in the analysis are provided in supplementary Table S1. The spatial coverage is shown in Fig. S1.

Fischer et al. (2022) compiled 31 different biome and land cover maps. We aggregated the maps from the 10×10 km spatial resolution to the 0.5° resolution of the trait data using the 'raster' R package (Hijmans, 2020). As the biome maps are categorical, we used the nearest neighbor method for aggregation. We removed biomes that occur in less than 40 grid cells and grid cells not covered by vegetation such as water and built-up areas.

2.2 Biome clustering

Following an approach previously applied by Boonman et al. (2022) and Scheiter et al. (2024), we used a supervised cluster
analysis to assign a biome type to each grid cell using trait information in the grid cell. Specifically, we used Gaussian mixture
models ('MclustDA' function in 'mclust' R package, Scrucca et al., 2016) and fitted models for each of the F31 biome maps.
The clustering of biomes was tested with different PFT-specific trait maps considering 1) only grasses, 2) only woody plants,
and 3) all PFTs (i.e., grasses and woody plants). In addition, we 4) combined traits of grasses and woody plants (i.e., combining
traits in 1 and 2). In case 4, the number of traits in the cluster analysis was twice the number of traits in cases 1, 2 and 3.

To assess the performance of the biome clustering, we used the κ statistics (Monserud and Leemans, 1992) to compare the
biome maps derived from the cluster analysis and the respective F31 map used to inform the clustering. The κ value quantifies
the agreement between categorical data sets. It considers the likelihood that agreement can occur by chance and is more robust
than calculating overlap. Values < 0 indicate no agreement, between 0 and 0.2 slight agreement, between 0.2 and 0.4 fair
agreement, between 0.4 and 0.6 moderate agreement, between 0.6 and 0.8 substantial agreement and between 0.8 and 1 almost
perfect agreement.

2.3 Trait ranking for classification

To test which traits are most suitable for biome classification, we used a randomized approach with a variable number of traits
and randomly sampled sets of traits. For each set of traits, cluster analyses were conducted for all of the F31 maps and for four
combinations of PFTs (see sec. 2.1). The number of traits in the clustering ranged between 3 and 12. Hence, cluster analyses
for grasses, woody plants and all PFTs included 3 to 12 variables whereas analyses with grass and woody plant traits combined
included 4, 6, ..., 24 variables. We repeated the clustering until at least 600 models were available for each F31 biome map
(21,867 models in total, Table S2). The performance of the clustering and its response to the number of traits included was
quantified using the κ statistics.

To identify the traits that are most suitable for biome classification, we ranked the traits following an approach previously
used by Scheiter et al. (2024). We selected all models with substantial agreement (i.e., $\kappa > 0.6$) from the randomized sensitivity
analysis with traits of grasses and woody plants combined. For the subset of models, we counted how often each trait was
included and expressed this number as percent value relative to the number of models. We interpret this value as a measure of
the suitability of traits for classification and ranked traits according to this measure.

We used a similar approach to rank the traits separately for each number of traits considered in the clustering while aggre-
gating models for all F31 maps. The κ values did not exceed 0.6 for low numbers of traits (see results). We therefore calculated
the 90% percentile of the κ values for each number of traits, and used only models with κ greater than the 90% percentile to
derive the ranking.

2.4 Specific trait combinations

To assess how data-model agreement was related to the F31 biome maps and the PFTs used for the clustering, the cluster
120 analyses were repeated for selected combinations of traits. (Cluster 1) We selected 6 traits with high rank in the sensitivity
analysis including all F31 biome maps and all numbers of traits (sec. 2.3). To account for traits describing different plant
components and those commonly used in the literature, and to avoid inclusion of similar traits (e.g., several traits describing
leaf nitrogen), we selected wood density, rooting depth, SLA, height, isotopic leaf nitrogen and conduit density (also see results
section). Then, we conducted a principal component analysis (PCA) using all traits to select traits that vary along different PCA
125 axes in trait space. Prior to the PCA, traits were normalized. (Cluster 2) We selected the three traits with highest loadings from
the first two PCA axes, these were SLA, leaf P, height, seed length, leaf dry matter content and single leaf mass. (Cluster 3) We
selected the two traits with highest loadings from the first three PCA axes, these were SLA, seed length, leaf dry matter content,
single leaf mass, seed germination rate and vessel length. (Cluster 4) Clustering was conducted for wood density, height and
SLA as in Boonman et al. (2022).

2.5 Biome-specific traits and model performance

Biome types used in different F31 biome maps differ substantially between maps. Therefore, biome-specific mean trait values
and κ values cannot be compared directly across all biome maps. We therefore calculated those values by selecting all biomes
across the F31 maps that share similarity in their names. Specifically, we merged all biome names of the F31 biome maps
as provided in the supplementary materials of Fischer et al. (2022) in one text string. Then we counted the occurrences of
135 all words in this string. After removing unnecessary words such as 'and' or 'with', we obtained a list of attributes defining
different biomes such as 'forest', 'evergreen' or 'boreal'. For the most frequent attributes (Table S3), we selected all biomes
from all F31 maps that contained this attribute in its name and calculated the mean trait values and κ for each of those biomes.
To calculate biome-specific κ values, we transformed the observation-based and the modeled biome maps into a binary map
where one represents the target biome and zero all other biomes. Then, we calculated κ for these binary maps.

140 This analysis showed that 'forest', 'tropical' and 'temperate' were used most frequently in the biome names with 179,
99 and 84 occurrences. We repeated the counting procedure separately for biomes containing each of those three attributes.
This procedure provided combinations such as 'evergreen forest' or 'tropical savanna'. For the most frequent combinations
of attributes (Table S4), we calculated the biome-specific mean trait values and κ values. Note, that this approach ignores
biome maps that do not use any of these attributes to denote biomes such as the Higgins et al. (2016) map, or the Netzel and
145 Stepinski (2016) map that denotes different biomes as numbered clusters. This analysis was conducted only for trait cluster 1,
as it showed the highest data-model agreement when κ values for all F31 biomes are averaged.

2.6 Global biome maps from traits

As the trait data derived from combining GBIF and TRY does not cover the entire global land surface, we extrapolated the
biome maps by using species distribution models (SDMs, Franklin, 2009). We extracted bioclimatic variables from WorldClim

150 (Hijmans, 2020). For the SDMs, we selected a subset of those variables: mean annual temperature (bio1), mean annual precip-
itation (bio12) and further uncorrelated variables (correlation<0.6, variables bio2, bio7, bio10, bio14, bio15, bio18). For each
biome of the considered biome map, an ensemble of 36 models was fitted by using GLM, CTA, ANN, SRE, FDA and RF,
three different sets of 3500 randomly selected pseudo-absences and two replicates. Then, an ensemble was derived by includ-
ing all models with True Skill Statistic TSS>0.5 and by combining models using the mean. By conducting predictions with

155 the ensemble model and the 10-minute WorldClim data, we derived global suitability maps for the different biomes. Finally,
suitability maps were aggregated to biome maps by identifying for each grid cell the biome with the highest suitability value.
The resulting biome map was compared to the corresponding observation-based biome map using the κ statistics and TSS.
Niche models were fitted using the 'biomod2' R package (Thuiller et al., 2023).

3 Results

3.1 Trait ranking for biome classification

The randomized sensitivity analysis showed that both the median and the maximum κ value increased when the number of
traits included in the cluster analysis increased (Fig. 1). The maximum κ value was 0.64 for clustering with 11 traits for grasses
and woody plants (22 traits in total). The median κ values saturated for higher numbers of traits, although the maximum value
increased further. For a given number of traits, maps based on grass traits showed the lowest predictive performance. Higher

165 model performance was obtained with trait maps derived from woody plants, all PFTs and grass and woody plants combined
(Figs. 1, S2).

When selecting cluster analyses with substantial agreement ($\kappa > 0.6$), leaf carbon, isotopic leaf nitrogen and SLA had the
highest rank, i.e., they were included in most models (in 41.1%, 40.3%, and 40.3% of the models, Fig. 2). Traits related to
seeds, leaf area and leaf mass had the lowest rank and were included in less than 30% of the models, seed number in only

170 12.6%. Despite the considerable variation among the relevance of the traits, we did not observe that traits are overly suitable
or unsuitable for predicting biomes. Conduit density, isotopic leaf nitrogen, leaf nitrogen per area and SLA were among the
highest-ranked traits for each number of traits in the sensitivity analysis (Fig. 3). Seed number had the lowest rank for all cases.

3.2 Specific trait clusters

Data-model agreement varied considerably for different biome maps and trait clusters used for the clustering (Fig. 4). When

175 averaged for all maps (averages of columns in Fig. 4), cluster analyses using traits of grass and woody plants combined showed
higher κ values than cluster analyses using only grasses, only woody plants or both PFTs. Using only traits of grasses showed
the lowest performance for all trait clusters. The highest average performance was obtained for cluster 1 (wood density, rooting
depth, SLA, height, isotopic leaf nitrogen and conduit density) with traits of grasses and woody plants combined ($\kappa = 0.44$).
When averaging all trait clusters for each biome map (averages of rows in Fig. 4), the highest performance was achieved for

180   the Nature Conservancy (2009) biome map ($\kappa = 0.45$) and the lowest performance for the Tateishi et al. (2011, 2014) map ($\kappa = 0.19$).

When considering different trait clusters and biome maps individually, the κ value was maximized when clustering was performed for the Olson et al. (2001) biome map with grass and woody plant traits of cluster 1 (κ=0.572), followed by the Nature Conservancy (2009) map (κ=0.569). Observation-based biomes were best predicted in Europe, North America, parts

185   of the Sahel region and Australia (Fig. 5). The biome types disagreed primarily in the subtropics. This result is also reflected in the biome-specific κ values of the clustering across the F31 maps (Fig. 6, S3-S6), where κ values were on average lower in tropical forests than in temperate or boreal forests.

For cluster 2 (SLA, leaf P, height, seed length, leaf dry matter content and single leaf mass), cluster 3 (SLA, seed length, leaf dry matter content, single leaf mass, seed germination rate and vessel length) and cluster 4 (SLA, wood density and height),

190   performance was maximized when using traits of grasses and woody plants separately and the Nature Conservancy (2009) biome map. The κ values were 0.545, 0.546 and 0.474, respectively.

3.3   Biome-specific traits

When selecting all biomes from the F31 maps that share similar attributes in their names, we found that mean trait values differ between biome types (Figs. 7, S7-S10). For example, tundra vegetation was characterized by shallow roots, low wood density

195   and high conduit density whereas savannas had deep roots and lower conduit density.

The attribute 'forest' was included in the names of 179 biomes in the F31 maps, and traits showed large variation across different forest types (Fig. S8). Splitting forests according to additional attributes revealed differences of mean trait values between forest types (Fig. 7). For example, tropical and subtropical forest showed deepest roots whereas needleleaf forest and boreal forest showed shallow roots. Deep or shallow roots co-occurred with lower or higher SLA and conduit density,

200   respectively. Similarly, splitting tropical and temperate biomes according to additional attributes revealed differences (Figs. S9, S10).

3.4   Predicting global biome patterns

Using SDMs, global biome maps can be predicted based on biome patterns derived from cluster analyses using trait data (Fig. 8). Similarly to biome patterns modeled for the spatial extent of the trait data, data-model agreement between the global

205   maps and the corresponding F31 map strongly differed. For example, for the Nature Conservancy (2009) map, $\kappa = 0.70$ and for the Tateishi et al. (2011, 2014) map, $\kappa = 0.19$.

4   Discussion

We used trait maps derived from combining crowd-sourced species distribution data and plant trait data to test if traits can be used to delineate global biomes. Although the coverage of trait data is heterogeneous and sparse in the tropics and sub-tropics,

210   and in the high northern latitudes of Canada and Siberia, the analyses showed that such a biome classification is possible. Yet,

the agreement between modeled and observation-based biome maps strongly depends on the traits and the biome map used to develop biome classification schemes.

4.1 Suitability of traits for biome classification

The performance of biome classification using traits was strongly related to the number and the selection of traits used for the
215 clustering. Previous studies on trait relationships (Wright et al., 2004; Díaz et al., 2016; Bruelheide et al., 2018) and biome classification using traits (van Bodegom et al., 2014; Boonman et al., 2022) focused on a limited number of traits and had a strong focus on leaf traits. Our results indicate that data-model agreement increases with the number of traits included in the analysis. This is not surprising given that additional traits add information to the clustering and represent additional trade-off axes in the trait space. Saturation of the performance for an increasing number of traits included indicates that the major trade-
220 off axes between traits that characterize biomes can be captured with around 10 traits or more. Saturation occurs because the traits included in the analysis are correlated (Fig. S11) and hence replaceable: when removing a trait from the analyses, its information may still be represented by another, correlated trait. Yet, selecting unsuitable traits lead to poor performance even if a high number of traits was used.

According to our analysis, height and leaf traits such as SLA, leaf carbon and leaf nitrogen were included in a high pro-
225 portion of the models with the highest performance and can be considered as most suitable for biome classification. This finding confirms the existence of generic plant strategies (Grime, 1988; Pierce et al., 2017) and the spectrum of plant form and function, where size and leaf economic traits were identified as the two major axes of the trait space (Díaz et al., 2016). The results highlight the importance of wood and root traits, particularly conduit density and rooting depth. Root traits gained more attention recently (Chave et al., 2009). They are often linked to plant hydraulics and water availability (Anderegg et al., 2018),
230 which makes them particularly important to understand how the projected increase of droughts under future climate change (IPCC, 2021) may influence ecosystem dynamics and resilience. Yet, measuring rooting traits is generally more difficult than measuring above-ground traits, which constrains the geographic coverage and taxonomic representativeness of root observations. According to our ranking, reproductive traits, particularly seed number, were least appropriate for biome classification. We attribute this result to the low coverage of seed traits in the TRY database, resulting in lower agreement between trait
235 data derived by merging TRY and GBIF data with independent sPlot trait data (Wolf et al., 2022). Seed number may also be determined by the dispersal mechanism of plants, a trait not considered in our analysis. Dispersal mechanisms are influenced for example by small-scale heterogeneity of landscapes, microclimate or biotic interaction networks (Schleuning et al., 2016).

We conducted a similar sensitivity analysis with traits simulated by the dynamic vegetation model aDGVM2 (Scheiter et al., 2024). This previous study showed that traits related to plant size were included in most models with high performance, while
240 leaf economic traits including SLA were less relevant. This contrasts the results of the study using observation-based trait data. Several reasons may explain this mismatch. First, some traits used in the aDGVM2 are not available in observation-based data sets, and vice versa. Traits such as isotopic leaf nitrogen or conduit density that turned out to have a high rank in the observation-based study are not implemented in the aDGVM2 (Langan et al., 2017; Kumar et al., 2021). Further, height and other size-related traits showed clear spatial patterns in response to environmental conditions in the aDGVM2, while the

245 patterns of other traits were less clear. Therefore, different biomes can be matched with different size classes simulated by the aDGVM2. Finally, in Scheiter et al. (2024) we focused on Africa, tropical Asia and Oceania. Data coverage in TRY is low in some of these regions while it is high in northern hemisphere biomes. This may imply that the ranking of traits is biased towards the tropics or the northern hemisphere in the modelling or the observation-based study, respectively.

4.2 Suitability of different biome maps to inform cluster analyses

250 Using different trait clusters for biome classification revealed consistent differences of data-model agreement for different biome maps used to inform the clustering. Overall, the Nature Conservancy (2009) map was best reproduced by the cluster analyses while the Tateishi et al. (2011, 2014) map showed the lowest performance. The maps assembled by Fischer et al. (2022) were derived from different data sources, including biogeographic zonation, species distributions, bioclimatic variables or remote sensing products including optical reflectance or NDVI. The performance of the clustering was highest for biome

255 maps based on biogeographic zonation and species distribution data (e.g., Nature Conservancy, 2009; Olson et al., 2001; Dinerstein et al., 2017) and lowest for biome maps based on reflectance (e.g., Tateishi et al., 2011, 2014). This can be explained by data sources in our study: we used species distribution data provided by the GBIF data base and trait information from TRY. These sources resemble those of biome maps using species distribution data, in contrast to biome maps based on reflectance. Cluster analyses including only traits with a direct imprint on Earth observation signals, such as optical reflectance (Cherif

260 et al., 2023; Kothari et al., 2023; Aguirre-Gutiérrez et al., 2021) may show higher agreement with such biome maps. Further, the remote sensing perspective from optical Earth observation satellites primarily informs on the top-canopy, while a representative biome characterization may require information on entire plant communities, including the understory (Dechant et al., 2023).

The systematic relation between data-model agreement and the biome maps used for the clustering was not found in the analysis using aDGVM2 results (Scheiter et al., 2024). In the previous study, performance was highest for the Beck et al. (2018)

265 map which was developed using bioclimatic variables. In the aDGVM2, simulated vegetation patterns were mostly defined by bioclimatic variables, that drive the major ecophysiological processes (Kumar et al., 2021). Hence, simulated vegetation closely follows bioclimatic gradients. Processes such as fire or competition can influence community composition within simulated grid cells but often not the large-scale patterns (Scheiter et al., 2024). The performance was lower in Scheiter et al. (2024, maximum value of $\kappa = 0.55$) than in analyses using observation-based data (maximum value of $\kappa = 0.64$).

270 According to our analysis, different biomes can be distinguished across the F31 maps based on trait values and co-occurring traits. For instance, boreal forests showed lower vegetation height, rooting depth and wood density, and higher SLA and conduit density than tropical forests. Yet, for other forest types, biome-specific trait means overlapped, such that the classification for these biomes is ambiguous. These results suggest, that the separation between different forest types is essential for accurate classification of the global biomes. To delineate biomes, we cannot rely on a few traits but need to consider the co-variation of

275 multiple traits in a multivariate trait space and optimization towards multiple ecological functions.

4.3 Recommendations for trait-based biome classification

We showed that using trait data for biome classification is feasible, but that several decisions regarding data and methods are necessary. These decisions are related to the trait data, the traits included in the analysis, the biome map used to develop a classification, and the method used for clustering.

280   Multiple methods have been applied to extrapolate traits from the site level to the global scale. The resulting trait maps can differ substantially as each data set is affected by different biases and uncertainties (Dechant et al., 2023; Wolf et al., 2022; Schiller et al., 2021). Accordingly, biome maps derived from different trait maps can differ. We therefore advise selecting trait data with high accuracy for biome classification and developing biome classification schemes for specific applications or regions using available and appropriate trait data.

285   Both the number of traits and the selection of traits influenced performance. Performance saturated as the number of traits increased. Yet, the selection of trait matters and including many unsuitable traits can imply low performance. Our results suggest that traits in biome classification should include leaf, stem and root traits as well as traits related to size to reflect different trade-off axes in the trait space. Ideally, the selected traits are only weakly correlated to avoid redundancy of information in the clustering. Yet, focusing only on weakly correlated traits can lead to poor performance if they do not reflect the global 290   spectrum of plant form and function (result not shown).

    When using traits for biome classification, data-model agreement was better when clustering was informed by biome maps based on vegetation data and biogeographic zonation than for maps based on remote sensing data. We therefore recommend using vegetation-based biome maps in the context of trait-based biome classification, such as Olson et al. (2001), Nature Conservancy (2009) or Dinerstein et al. (2017).

295   Here, we used Gaussian mixture models for the cluster analysis. Multiple alternative approaches are available, both for supervised and unsupervised classification. A systematic assessment of alternative approaches has not been performed, but may be valuable to identify the most appropriate method. Unsupervised classification can be used to create novel biome maps that are only defined by trait information but not by existing biome or land-cover maps. Alternatively, dimensionality reduction methods such as principal component analysis or isometric feature mapping could be applied to explore a trait-based, 300   continuous representation of biome patterns.

5 Conclusions

The crowd-sourced trait data utilized for the analysis was spatially heterogeneous with large gaps in parts of the tropics, subtropics, and high northern latitudes. Nonetheless, the data were suitable for delimiting global biomes using cluster analyses and to predict global biome maps. This result highlights the value of crowd-sourced trait- and species distribution data in 305   the biogeography of biomes, despite the data gaps. We argue that filling gaps in the trait data would not only yield a more comprehensive understanding of the spectrum of plant form and function, but also allow more accurate biome classification. We also showed that increasing the number of traits in the cluster analysis improved model performance, highlighting the need to fill data gaps with respect to traits available for specific species or sites. Finally, our results can inform the development

of dynamic vegetation models as it showed which traits are important for biome classification and which trait values are
310 characteristic for different biomes. While leaf economic traits are already well-captured in such models, we argue that the
representation of wood and root traits should be improved.

Code and data availability. Code and data that support the findings of this study are openly available.
1) Trait data: https://github.com/tejakattenborn/GBIF_trait_maps/
2) Bioclimatic data for species distribution models: https://www.worldclim.org
315 3) Fischer et al. (2022) biome maps: https://doi.org/10.5061/dryad.hqbzkh1jm.
4) All R scripts required to conduct the analyses and generate plots are available in Zenodo: https://zenodo.org/records/10526277. Access is
currently restricted but will be granted to editors and referees on request. The repository will be made open access if manuscript is accepted
for publication.

Author contributions. SW and TK created the trait data; SS conceived the study, performed the analyses and wrote the first draft of the
320 manuscript; SW and TK contributed to the analyses and writing of the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. TK ancknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, Emmy Noether project PANOPS, grant number 504978936).

References

325   Aguirre-Gutiérrez, J., Rifai, S., Shenkin, A., Oliveras, I., Bentley, L. P., Svátek, M., Girardin, C. A. J., Both, S., Riutta, T., Berenguer, E., Kissling, W. D., Bauman, D., Raab, N., Moore, S., Farfan-Rios, W., Figueiredo, A. E. S., Reis, S. M., Ndong, J. E., Ondo, F. E., N'ssi Bengone, N., Mihindou, V., Moraes de Seixas, M. M., Adu-Bredu, S., Abernethy, K., Asner, G. P., Barlow, J., Burslem, D. F. R. P., Coomes, D. A., Cernusak, L. A., Dargie, G. C., Enquist, B. J., Ewers, R. M., Ferreira, J., Jeffery, K. J., Joly, C. A., Lewis, S. L., Marimon-Junior, B. H., Martin, R. E., Morandi, P. S., Phillips, O. L., Quesada, C. A., Salinas, N., Schwantes Marimon, B., Silman, M., Teh, Y. A.,
330   White, L. J. T., and Malhi, Y.: Pantropical modelling of canopy functional traits using Sentinel-2 remote sensing data, Remote Sensing of Environment, 252, 112 122, https://doi.org/10.1016/j.rse.2020.112122, 2021.

     Anderegg, W. R. L., Konings, A. G., Trugman, A. T., Yu, K., Bowling, D. R., Gabbitas, R., Karp, D. S., Pacala, S., Sperry, J. S., Sulman, B. N., and Zenes, N.: Hydraulic diversity of forests regulates ecosystem resilience during drought, Nature, 561, 538–541, https://doi.org/10.1038/s41586-018-0539-7, 2018.

335   Barnes, R. and Sahr, K.: dggridR: Discrete Global Grids, https://CRAN.R-project.org/package=dggridR, r package version 3.0.0, 2023.

     Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., and Wood, E. F.: Present and future Köppen-Geiger climate classification maps at 1-km resolution, Scientific Data, 5, 180 214, https://doi.org/10.1038/sdata.2018.214, 2018.

     Beierkuhnlein, C. and Fischer, J.-C.: Global biomes and ecozones – Conceptual and spatial communalities and discrepancies, Erdkunde, 75, 249–270, 2021.

340   Boonman, C. C. F., Benítez-López, A., Schipper, A. M., Thuiller, W., Anand, M., Cerabolini, B. E. L., Cornelissen, J. H. C., Gonzalez-Melo, A., Hattingh, W. N., Higuchi, P., Laughlin, D. C., Onipchenko, V. G., Peñuelas, J., Poorter, L., Soudzilovskaia, N. A., Huijbregts, M. A. J., and Santini, L.: Assessing the reliability of predicted plant trait distributions at the global scale, Global Ecology and Biogeography, 29, 1034–1051, https://doi.org/https://doi.org/10.1111/geb.13086, 2020.

     Boonman, C. C. F., Huijbregts, M. A. J., Benítez-López, A., Schipper, A. M., Thuiller, W., and Santini, L.: Trait-based projections of climate
345   change effects on global biome distributions, Diversity and Distributions, 28, 25–37, https://doi.org/https://doi.org/10.1111/ddi.13431, 2022.

     Bruelheide, H., Dengler, J., Purschke, O., Lenoir, J., Jiménez-Alfaro, B., Hennekens, S. M., Botta-Dukát, Z., Chytrý, M., Field, R., Jansen, F., Kattge, J., Pillar, V. D., Schrodt, F., Mahecha, M. D., Peet, R. K., Sandel, B., van Bodegom, P., Altman, J., Alvarez-Dávila, E., Arfin Khan, M. A. S., Attorre, F., Aubin, I., Baraloto, C., Barroso, J. G., Bauters, M., Bergmeier, E., Biurrun, I., Bjorkman, A. D., Blonder, B., Čarni,
350   A., Cayuela, L., Černý, T., Cornelissen, J. H. C., Craven, D., Dainese, M., Derroire, G., De Sanctis, M., Díaz, S., Doležal, J., Farfan-Rios, W., Feldpausch, T. R., Fenton, N. J., Garnier, E., Guerin, G. R., Gutiérrez, A. G., Haider, S., Hattab, T., Henry, G., Hérault, B., Higuchi, P., Hölzel, N., Homeier, J., Jentsch, A., Jürgens, N., Kącki, Z., Karger, D. N., Kessler, M., Kleyer, M., Knollová, I., Korolyuk, A. Y., Kühn, I., Laughlin, D. C., Lens, F., Loos, J., Louault, F., Lyubenova, M. I., Malhi, Y., Marcenò, C., Mencuccini, M., Müller, J. V., Munzinger, J., Myers-Smith, I. H., Neill, D. A., Niinemets, Ü., Orwin, K. H., Ozinga, W. A., Penuelas, J., Pérez-Haase, A., Petřík, P., Phillips, O. L.,
355   Pärtel, M., Reich, P. B., Römermann, C., Rodrigues, A. V., Sabatini, F. M., Sardans, J., Schmidt, M., Seidler, G., Silva Espejo, J. E., Silveira, M., Smyth, A., Sporbert, M., Svenning, J.-C., Tang, Z., Thomas, R., Tsiripidis, I., Vassilev, K., Violle, C., Virtanen, R., Weiher, E., Welk, E., Wesche, K., Winter, M., Wirth, C., and Jandt, U.: Global trait–environment relationships of plant communities, Nature Ecology & Evolution, 2, 1906–1917, https://doi.org/10.1038/s41559-018-0699-8, 2018.

     Chave, J., Coomes, D., Jansen, S., Lewis, S. L., Swenson, N. G., and Zanne, A. E.: Towards a worldwide wood economics spectrum, Ecology
360   Letters, 12, 351–366, 2009.

Cherif, E., Feilhauer, H., Berger, K., Dao, P. D., Ewald, M., Hank, T. B., He, Y., Kovach, K. R., Lu, B., Townsend, P. A., and Kattenborn, T.: From spectra to plant functional traits: Transferable multi-trait models from heterogeneous and sparse data, Remote Sensing of Environment, 292, 113 580, https://doi.org/10.1016/j.rse.2023.113580, 2023.

Dechant, B. et al.: Intercomparison of global foliar trait maps reveals fundamental differences and limitations of upscaling approaches, EarthArXiv, XX, XX, https://doi.org/10.31223/X58S97, 2023.

Díaz, S., Kattge, J., Cornelissen, J. H. C., Wright, I. J., Lavorel, S., Dray, S., Reu, B., Kleyer, M., Wirth, C., Colin Prentice, I., Garnier, E., Bönisch, G., Westoby, M., Poorter, H., Reich, P. B., Moles, A. T., Dickie, J., Gillison, A. N., Zanne, A. E., Chave, J., Joseph Wright, S., Sheremet'ev, S. N., Jactel, H., Baraloto, C., Cerabolini, B., Pierce, S., Shipley, B., Kirkup, D., Casanoves, F., Joswig, J. S., Günther, A., Falczuk, V., Rüger, N., Mahecha, M. D., and Gorné, L. D.: The global spectrum of plant form and function, Nature, 529, 167–171, https://doi.org/10.1038/nature16489, 2016.

Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N. D., Wikramanayake, E., Hahn, N., Palminteri, S., Hedao, P., Noss, R., Hansen, M., Locke, H., Ellis, E. C., Jones, B., Barber, C. V., Hayes, R., Kormos, C., Martin, V., Crist, E., Sechrest, W., Price, L., Baillie, J. E. M., Weeden, D., Suckling, K., Davis, C., Sizer, N., Moore, R., Thau, D., Birch, T., Potapov, P., Turubanova, S., Tyukavina, A., de Souza, N., Pintea, L., Brito, J. C., Llewellyn, O. A., Miller, A. G., Patzelt, A., Ghazanfar, S. A., Timberlake, J., Klöser, H., Shennan-Farpón, Y., Kindt, R., Lillesø, J.-P. B., van Breugel, P., Graudal, L., Voge, M., Al-Shammari, K. F., and Saleem, M.: An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm, BioScience, 67, 534–545, https://doi.org/10.1093/biosci/bix014, 2017.

Fischer, J.-C., Walentowitz, A., and Beierkuhnlein, C.: The biome inventory - standardizing global biogeographical land units, Global Ecology and Biogeography, 31, 2172–2183, https://doi.org/https://doi.org/10.1111/geb.13574, 2022.

Franklin, J.: Mapping species distributions. Spatial inference and prediction, Cambridge University Press, 2009.

GBIF.Org User: Occurrence Download, https://doi.org/10.15468/DL.FE2KV3, 2023.

Grime, J. P.: Plant evolutionary biology, chap. The C-S-R model of primary plant strategies - origins, implications and tests, pp. 371–393, Chapman and Hall, London and New York, 1988.

Higgins, S. I., Buitenwerf, R., and Moncrieff, G. R.: Defining functional biomes and monitoring their change globally, Global Change Biology, 22, 3583–3593, 2016.

Hijmans, R. J.: raster: Geographic Data Analysis and Modeling, https://CRAN.R-project.org/package=raster, r package version 3.0-12, 2020.

Hunter, J., Franklin, S., Luxton, S., and Loidi, J.: Terrestrial biomes: a conceptual review, VCS, 2, 73–85, https://doi.org/10.3897/VCS/2021/61463, 2021.

IPCC: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press. In Press, 2021.

Kattge, J., Bönisch, G., Díaz, S., Lavorel, S., Prentice, I. C., Leadley, P., Tautenhahn, S., Werner, G. D. A., Aakala, T., Abedi, M., Acosta, A. T. R., Adamidis, G. C., Adamson, K., Aiba, M., Albert, C. H., Alcántara, J. M., Alcázar C, C., Aleixo, I., Ali, H., Amiaud, B., Ammer, C., Amoroso, M. M., Anand, M., Anderson, C., Anten, N., Antos, J., Apgaua, D. M. G., Ashman, T.-L., Asmara, D. H., Asner, G. P., Aspinwall, M., Atkin, O., Aubin, I., Baastrup-Spohr, L., Bahalkeh, K., Bahn, M., Baker, T., Baker, W. J., Bakker, J. P., Baldocchi, D., Baltzer, J., Banerjee, A., Baranger, A., Barlow, J., Barneche, D. R., Baruch, Z., Bastianelli, D., Battles, J., Bauerle, W., Bauters, M., Bazzato, E., Beckmann, M., Beeckman, H., Beierkuhnlein, C., Bekker, R., Belfry, G., Belluau, M., Beloiu, M., Benavides, R., Benomar, L., Berdugo-Lattke, M. L., Berenguer, E., Bergamin, R., Bergmann, J., Bergmann Carlucci, M., Berner, L., Bernhardt-Römermann, M., Bigler, C., Bjorkman, A. D., Blackman, C., Blanco, C., Blonder, B., Blumenthal, D., Bocanegra-González, K. T., Boeckx, P., Bohlman, S., Böhning-Gaese, K., Boisvert-Marsh, L., Bond, W., Bond-Lamberty, B., Boom, A., Boonman, C. C. F., Bordin, K., Boughton, E. H.,

Boukili, V., Bowman, D. M. J. S., Bravo, S., Brendel, M. R., Broadley, M. R., Brown, K. A., Bruelheide, H., Brumnich, F., Bruun,
400 H. H., Bruy, D., Buchanan, S. W., Bucher, S. F., Buchmann, N., Buitenwerf, R., Bunker, D. E., Bürger, J., Burrascano, S., Burslem, D.
F. R. P., Butterfield, B. J., Byun, C., Marques, M., Scalon, M. C., Caccianiga, M., Cadotte, M., Cailleret, M., Camac, J., Camarero, J. J.,
Campany, C., Campetella, G., Campos, J. A., Cano-Arboleda, L., Canullo, R., Carbognani, M., Carvalho, F., Casanoves, F., Castagneyrol,
B., Catford, J. A., Cavender-Bares, J., Cerabolini, B. E. L., Cervellini, M., Chacón-Madrigal, E., Chapin, K., Chapin, F. S., Chelli, S.,
Chen, S.-C., Chen, A., Cherubini, P., Chianucci, F., Choat, B., Chung, K.-S., Chytrý, M., Ciccarelli, D., Coll, L., Collins, C. G., Conti, L.,
405 Coomes, D., Cornelissen, J. H. C., Cornwell, W. K., Corona, P., Coyea, M., Craine, J., Craven, D., Cromsigt, J. P. G. M., Csecserits, A.,
Cufar, K., Cuntz, M., da Silva, A. C., Dahlin, K. M., Dainese, M., Dalke, I., Dalle Fratte, M., Dang-Le, A. T., Danihelka, J., Dannoura,
M., Dawson, S., de Beer, A. J., De Frutos, A., De Long, J. R., Dechant, B., Delagrange, S., Delpierre, N., Derroire, G., Dias, A. S.,
Diaz-Toribio, M. H., Dimitrakopoulos, P. G., Dobrowolski, M., Doktor, D., Dřevojan, P., Dong, N., Dransfield, J., Dressler, S., Duarte,
L., Ducouret, E., Dullinger, S., Durka, W., Duursma, R., Dymova, O., E-Vojtkó, A., Eckstein, R. L., Ejtehadi, H., Elser, J., Emilio, T.,
410 Engemann, K., Erfanian, M. B., Erfmeier, A., Esquivel-Muelbert, A., Esser, G., Estiarte, M., Domingues, T. F., Fagan, W. F., Fagúndez, J.,
Falster, D. S., Fan, Y., Fang, J., Farris, E., Fazlioglu, F., Feng, Y., Fernandez-Mendez, F., Ferrara, C., Ferreira, J., Fidelis, A., Finegan, B.,
Firn, J., Flowers, T. J., Flynn, D. F. B., Fontana, V., Forey, E., Forgiarini, C., François, L., Frangipani, M., Frank, D., Frenette-Dussault,
C., Freschet, G. T., Fry, E. L., Fyllas, N. M., Mazzochini, G. G., Gachet, S., Gallagher, R., Ganade, G., Ganga, F., García-Palacios, P.,
Gargaglione, V., Garnier, E., Garrido, J. L., de Gasper, A., Gea-Izquierdo, G., Gibson, D., Gillison, A. N., Giroldo, A., Glasenhardt, M.-C.,
415 Gleason, S., Gliesch, M., Goldberg, E., Göldel, B., Gonzalez-Akre, E., Gonzalez-Andujar, J. L., González-Melo, A., González-Robles,
A., Graae, B. J., Granda, E., Graves, S., Green, W. A., Gregor, T., Gross, N., Guerin, G. R., Günther, A., Gutiérrez, A. G., Haddock, L.,
Haines, A., Hall, J., Hambuckers, A., Han, W., Harrison, S. P., Hattingh, W., Hawes, J. E., He, T., He, P., Heberling, J. M., Helm, A.,
Hempel, S., Hentschel, J., Hérault, B., Hereş, A.-M., Herz, K., Heuertz, M., Hickler, T., Hietz, P., Higuchi, P., Hipp, A. L., Hirons, A.,
Hock, M., Hogan, J. A., Holl, K., Honnay, O., Hornstein, D., Hou, E., Hough-Snee, N., Hovstad, K. A., Ichie, T., Igić, B., Illa, E., Isaac,
420 M., Ishihara, M., Ivanov, L., Ivanova, L., Iversen, C. M., Izquierdo, J., Jackson, R. B., Jackson, B., Jactel, H., Jagodzinski, A. M., Jandt,
U., Jansen, S., Jenkins, T., Jentsch, A., Jespersen, J. R. P., Jiang, G.-F., Johansen, J. L., Johnson, D., Jokela, E. J., Joly, C. A., Jordan,
G. J., Joseph, G. S., Junaedi, D., Junker, R. R., Justes, E., Kabzems, R., Kane, J., Kaplan, Z., Kattenborn, T., Kavelenova, L., Kearsley,
E., Kempel, A., Kenzo, T., Kerkhoff, A., Khalil, M. I., Kinlock, N. L., Kissling, W. D., Kitajima, K., Kitzberger, T., Kjøller, R., Klein, T.,
Kleyer, M., Klimešová, J., Klipel, J., Kloeppel, B., Klotz, S., Knops, J. M. H., Kohyama, T., Koike, F., Kollmann, J., Komac, B., Komatsu,
425 K., König, C., Kraft, N. J. B., Kramer, K., Kreft, H., Kühn, I., Kumarathunge, D., Kuppler, J., Kurokawa, H., Kurosawa, Y., Kuyah, S.,
Laclau, J.-P., Lafleur, B., Lallai, E., Lamb, E., Lamprecht, A., Larkin, D. J., Laughlin, D., Le Bagousse-Pinguet, Y., le Maire, G., le Roux,
P. C., le Roux, E., Lee, T., Lens, F., Lewis, S. L., Lhotsky, B., Li, Y., Li, X., Lichstein, J. W., Liebergesell, M., Lim, J. Y., Lin, Y.-S.,
Linares, J. C., Liu, C., Liu, D., Liu, U., Livingstone, S., Llusià, J., Lohbeck, M., López-García, Á., Lopez-Gonzalez, G., Lososová, Z.,
Louault, F., Lukács, B. A., Lukeš, P., Luo, Y., Lussu, M., Ma, S., Maciel Rabelo Pereira, C., Mack, M., Maire, V., Mäkelä, A., Mäkinen,
430 H., Malhado, A. C. M., Mallik, A., Manning, P., Manzoni, S., Marchetti, Z., Marchino, L., Marcilio-Silva, V., Marcon, E., Marignani,
M., Markesteijn, L., Martin, A., Martínez-Garza, C., Martínez-Vilalta, J., Mašková, T., Mason, K., Mason, N., Massad, T. J., Masse, J.,
Mayrose, I., McCarthy, J., McCormack, M. L., McCulloh, K., McFadden, I. R., McGill, B. J., McPartland, M. Y., Medeiros, J. S., Medlyn,
B., Meerts, P., Mehrabi, Z., Meir, P., Melo, F. P. L., Mencuccini, M., Meredieu, C., Messier, J., Mészáros, I., Metsaranta, J., Michaletz,
S. T., Michelaki, C., Migalina, S., Milla, R., Miller, J. E. D., Minden, V., Ming, R., Mokany, K., Moles, A. T., Molnár V, A., Molofsky, J.,
435 Molz, M., Montgomery, R. A., Monty, A., Moravcová, L., Moreno-Martínez, A., Moretti, M., Mori, A. S., Mori, S., Morris, D., Morrison,
J., Mucina, L., Mueller, S., Muir, C. D., Müller, S. C., Munoz, F., Myers-Smith, I. H., Myster, R. W., Nagano, M., Naidu, S., Narayanan,

A., Natesan, B., Negoita, L., Nelson, A. S., Neuschulz, E. L., Ni, J., Niedrist, G., Nieto, J., Niinemets, Ü., Nolan, R., Nottebrock, H., Nouvellon, Y., Novakovskiy, A., Network, T. N., Nystuen, K. O., O'Grady, A., O'Hara, K., O'Reilly-Nugent, A., Oakley, S., Oberhuber, W., Ohtsuka, T., Oliveira, R., Öllerer, K., Olson, M. E., Onipchenko, V., Onoda, Y., Onstein, R. E., Ordonez, J. C., Osada, N., Ostonen,
440   I., Ottaviani, G., Otto, S., Overbeck, G. E., Ozinga, W. A., Pahl, A. T., Paine, C. E. T., Pakeman, R. J., Papageorgiou, A. C., Parfionova, E., Pärtel, M., Patacca, M., Paula, S., Paule, J., Pauli, H., Pausas, J. G., Peco, B., Penuelas, J., Perea, A., Peri, P. L., Petisco-Souza, A. C., Petraglia, A., Petritan, A. M., Phillips, O. L., Pierce, S., Pillar, V. D., Pisek, J., Pomogaybin, A., Poorter, H., Portsmuth, A., Poschlod, P., Potvin, C., Pounds, D., Powell, A. S., Power, S. A., Prinzing, A., Puglielli, G., Pyšek, P., Raevel, V., Rammig, A., Ransijn, J., Ray, C. A., Reich, P. B., Reichstein, M., Reid, D. E. B., Réjou-Méchain, M., de Dios, V. R., Ribeiro, S., Richardson, S., Riibak, K., Rillig, M. C.,
445   Riviera, F., Robert, E. M. R., Roberts, S., Robroek, B., Roddy, A., Rodrigues, A. V., Rogers, A., Rollinson, E., Rolo, V., Römermann, C., Ronzhina, D., Roscher, C., Rosell, J. A., Rosenfield, M. F., Rossi, C., Roy, D. B., Royer-Tardif, S., Rüger, N., Ruiz-Peinado, R., Rumpf, S. B., Rusch, G. M., Ryo, M., Sack, L., Saldaña, A., Salgado-Negret, B., Salguero-Gomez, R., Santa-Regina, I., Santacruz-García, A. C., Santos, J., Sardans, J., Schamp, B., Scherer-Lorenzen, M., Schleuning, M., Schmid, B., Schmidt, M., Schmitt, S., Schneider, J. V., Schowanek, S. D., Schrader, J., Schrodt, F., Schuldt, B., Schurr, F., Selaya Garvizu, G., Semchenko, M., Seymour, C., Sfair, J. C., Sharpe,
450   J. M., Sheppard, C. S., Sheremetiev, S., Shiodera, S., Shipley, B., Shovon, T. A., Siebenkäs, A., Sierra, C., Silva, V., Silva, M., Sitzia, T., Sjöman, H., Slot, M., Smith, N. G., Sodhi, D., Soltis, P., Soltis, D., Somers, B., Sonnier, G., Sørensen, M. V., Sosinski Jr, E. E., Soudzilovskaia, N. A., Souza, A. F., Spasojevic, M., Sperandii, M. G., Stan, A. B., Stegen, J., Steinbauer, K., Stephan, J. G., Sterck, F., Stojanovic, D. B., Strydom, T., Suarez, M. L., Svenning, J.-C., Svitková, I., Svitok, M., Svoboda, M., Swaine, E., Swenson, N., Tabarelli, M., Takagi, K., Tappeiner, U., Tarifa, R., Tauugourdeau, S., Tavsanoglu, C., te Beest, M., Tedersoo, L., Thiffault, N., Thom, D., Thomas,
455   E., Thompson, K., Thornton, P. E., Thuiller, W., Tichý, L., Tissue, D., Tjoelker, M. G., Tng, D. Y. P., Tobias, J., Török, P., Tarin, T., Torres-Ruiz, J., Tóthmérész, B., Treurnicht, M., Trivellone, V., Trolliet, F., Trotsiuk, V., Tsakalos, J. L., Tsiripidis, I., Tysklind, N., Umehara, T., Usoltsev, V., Vadeboncoeur, M., Vaezi, J., Valladares, F., Vamosi, J., van Bodegom, P. M., van Breugel, M., Van Cleemput, E., van de Weg, M., van der Merwe, S., van der Plas, F., van der Sande, M. T., van Kleunen, M., Van Meerbeek, K., Vanderwel, M., Vanselow, K. A., Vårhammar, A., Varone, L., Vasquez Valderrama, M. Y., Vassilev, K., Vellend, M., Veneklaas, E. J., Verbeeck, H., Verheyen, K., Vibrans,
460   A., Vieira, I., Villacís, J., Violle, C., Vivek, P., Wagner, K., Waldram, M., Waldron, A., Walker, A. P., Waller, M., Walther, G., Wang, H., Wang, F., Wang, W., Watkins, H., Watkins, J., Weber, U., Weedon, J. T., Wei, L., Weigelt, P., Weiher, E., Wells, A. W., Wellstein, C., Wenk, E., Westoby, M., Westwood, A., White, P. J., Whitten, M., Williams, M., Winkler, D. E., Winter, K., Womack, C., Wright, I. J., Wright, S. J., Wright, J., Pinho, B. X., Ximenes, F., Yamada, T., Yamaji, K., Yanai, R., Yankov, N., Yguel, B., Zanini, K. J., Zanne, A. E., Zelený, D., Zhao, Y.-P., Zheng, J., Zheng, J., Ziemińska, K., Zirbel, C. R., Zizka, G., Zo-Bi, I., Zotz, G., and Wirth, C.: TRY plant trait
465   database - enhanced coverage and open access, Global Change Biology, 26, 119–188, https://doi.org/10.1111/gcb.14904, 2020.

Kothari, S., Beauchamp-Rioux, R., Blanchard, F., Crofts, A. L., Girard, A., Guilbeault-Mayers, X., Hacker, P. W., Pardo, J., Schweiger, A. K., Demers-Thibeault, S., Bruneau, A., Coops, N. C., Kalacska, M., Vellend, M., and Laliberté, E.: Predicting leaf traits across functional groups using reflectance spectroscopy, New Phytologist, 238, 549–566, https://doi.org/10.1111/nph.18713, 2023.

Kumar, D., Pfeiffer, M., Gaillard, C., Langan, L., and Scheiter, S.: Climate change and elevated CO_2 favor forest over savanna under different
470   future scenarios in South Asia, Biogeoscience, 18, 2957–2979, https://doi.org/10.5194/bg-18-2957-2021, 2021.

Langan, L., Higgins, S. I., and Scheiter, S.: Climate-biomes, pedo-biomes or pyro-biomes: which world view explains the tropical forest – savanna boundary in South America?, Journal of Biogeography, 44, 2319–2330, 2017.

Ludwig, M., Moreno-Martinez, A., Hölzel, N., Pebesma, E., and Meyer, H.: Assessing and improving the transferability of current global spatial prediction models, Global Ecology and Biogeography, 32, 356–368, https://doi.org/10.1111/geb.13635, 2023.

475   Monserud, R. A. and Leemans, R.: Comparing global vegetation maps with the kappa-statistic, Ecological Modelling, 62, 275–293, 1992.

Mucina, L.: Biome: evolution of a crucial ecological and biogeographical concept, New Phytologist, 222, 97–114, https://doi.org/10.1111/nph.15609, 2019.

Nature Conservancy: Global ecoregions, major habitat types, biogeographical realms and the nature conservancy terrestrial assessment units as of December 14, 2009, Tech. rep., The Nature Conservancy, 2009.

480   Netzel, P. and Stepinski, T.: On using a clustering approach for global climate classification, Journal of Climate, 29, 3387–3401, 2016.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., D'amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P., and Kassem, K. R.: Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity, BioScience, 51, 933–938, https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2, 2001.

485   Pierce, S., Negreiros, D., Cerabolini, B. E. L., Kattge, J., Díaz, S., Kleyer, M., Shipley, B., Wright, S. J., Soudzilovskaia, N. A., Onipchenko, V. G., van Bodegom, P. M., Frenette-Dussault, C., Weiher, E., Pinho, B. X., Cornelissen, J. H. C., Grime, J. P., Thompson, K., Hunt, R., Wilson, P. J., Buffa, G., Nyakunga, O. C., Reich, P. B., Caccianiga, M., Mangili, F., Ceriani, R. M., Luzzaro, A., Brusa, G., Siefert, A., Barbosa, N. P. U., Chapin III, F. S., Cornwell, W. K., Fang, J., Fernandes, G. W., Garnier, E., Le Stradic, S., Peñuelas, J., Melo, F. P. L., Slaviero, A., Tabarelli, M., and Tampucci, D.: A global method for calculating plant CSR ecological strategies applied across biomes
490   world-wide, Functional Ecology, 31, 444–457, https://doi.org/10.1111/1365-2435.12722, 2017.

Scheiter, S., Kumar, D., Pfeiffer, M., and Langan, L.: Biome classification influences current and projected future biome distributions, Global Ecology and Biogeography, 33, 259–271, 2024.

Schiller, C., Schmidtlein, S., Boonman, C., Moreno-Martínez, A., and Kattenborn, T.: Deep learning and citizen science enable automated plant trait predictions from photographs, Scientific Reports, 11, 16 395, 2021.

495   Schleuning, M., Fründ, J., Schweiger, O., Welk, E., Albrecht, J., Albrecht, M., Beil, M., Benadi, G., Blüthgen, N., Bruelheide, H., Böhning-Gaese, K., Dehling, D. M., Dormann, C. F., Exeler, N., Farwig, N., Harpke, A., Hickler, T., Kratochwil, A., Kuhlmann, M., Kühn, I., Michez, D., Mudri-Stojnić, S., Plein, M., Rasmont, P., Schwabe, A., Settele, J., Vujić, A., Weiner, C. N., Wiemers, M., and Hof, C.: Ecological networks are more sensitive to plant than to animal extinction under climate change, Nature Communications, 7, 13 965, https://doi.org/10.1038/ncomms13965, 2016.

500   Schrodt, F., Kattge, J., Shan, H., Fazayeli, F., Joswig, J., Banerjee, A., Reichstein, M., Bönisch, G., Díaz, S., Dickie, J., et al.: BHPMF–a hierarchical Bayesian approach to gap-filling and trait prediction for macroecology and functional biogeography, Global Ecology and Biogeography, 24, 1510–1521, 2015.

Scrucca, L., Fop, M., Murphy, T. B., and Raftery, A. E.: mclust5: clustering, classification and density estimation using Gaussian finite mixture models, The R Journal, 8, 289–317, 2016.

505   Tateishi, R., Bayaer, U., Al-Bilbisi, H., Aboel Ghar, M., Tsend-Ayush, J., Kobayashi, T., Kasimu, A., Hoan, N. T., Shalaby, A., Alsaaideh, B., Enkhzaya, T., G., and Sato, H. P.: Production of global land cover data - GLCNMO, International Journal of Digital Earth, 4, 22–49, https://doi.org/10.1080/17538941003777521, 2011.

Tateishi, R., Thanh Hoan, N., Kobayashi, T., Alsaaideh, B., Tana, G., and Xuan Phong, D.: Production of global land cover data - GLC-NMO2008, Journal of Geography and Geology, 6, 99–122, 2014.

510   Thuiller, W., Georges, D., Gueguen, M., Engler, R., Breiner, F., Lafourcade, B., and Patin, R.: biomod2: Ensemble Platform for Species Distribution Modeling, https://CRAN.R-project.org/package=biomod2, r package version 4.2-4, 2023.

van Bodegom, P. M., Douma, J. C., and Verheijen, L. M.: A fully traits-based approach to modeling global vegetation distribution, Proceedings of the National Academy of Sciences, 111, 13 733–13 738, https://doi.org/10.1073/pnas.1304551110, 2014.

Wolf, S., Mahecha, M. D., Sabatini, F. M., Wirth, C., Bruelheide, H., Kattge, J., Moreno Martínez, Á., Mora, K., and Kattenborn, T.: Citizen
515   science plant observations encode global trait patterns, Nature Ecology & Evolution, https://doi.org/10.1038/s41559-022-01904-x, 2022.

Wright, I. J., Reich, P. B., Westoby, M., Ackerly, D. D., Baruch, Z., Bongers, F., Cavender-Bares, J., Chapin, T., Cornelissen, J. H. C., Diemer, M., Flexas, J., Garnier, E., Groom, P. K., Gulias, J., Hikosaka, K., Lamont, B. B., Lee, T., Lee, W., Lusk, C., Midgley, J. J., Navas, M. L., Niinemets, U., Oleksyn, J., Osada, N., Poorter, H., Poot, P., Prior, L., Pyankov, V. I., Roumet, C., Thomas, S. C., Tjoelker, M. G., Veneklaas, E. J., and Villar, R.: The worldwide leaf economics spectrum, Nature, 428, 821–827, 2004.

520   Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Duarte Ritter, C., Edler, D., Farooq, H., Herdean, A., Ariza, M., Scharn, R., Svanteson, S., Wengstrom, N., Zizka, V., and Antonelli, A.: CoordinateCleaner: standardized cleaning of occurrence records from biological collection databases, Methods in Ecology and Evolution, pp. –7, https://doi.org/10.1111/2041-210X.13152, r package version 3.0.1, 2019.
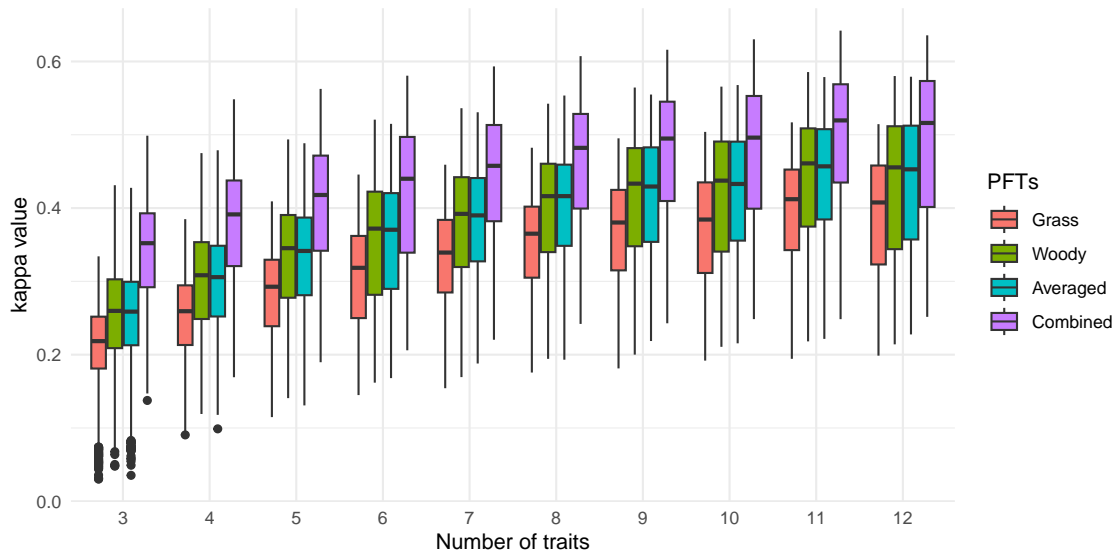
Figure 1. Relation between data-model agreement and the number of traits included in the cluster analysis. For each number of traits, the traits were randomly selected, and clustering was conducted for all F31 biome maps. At least 600 cluster analysis were conducted for each F31 map. Traits were sampled from those provided in Table S1 for different combinations of PFTs. Data-model agreement is represented by the κ statistics. Note that 'Combined' indicates clustering with traits of both grasses and woody plants, such that the number of traits is twice the number provided in the figure. See also Fig. S2.
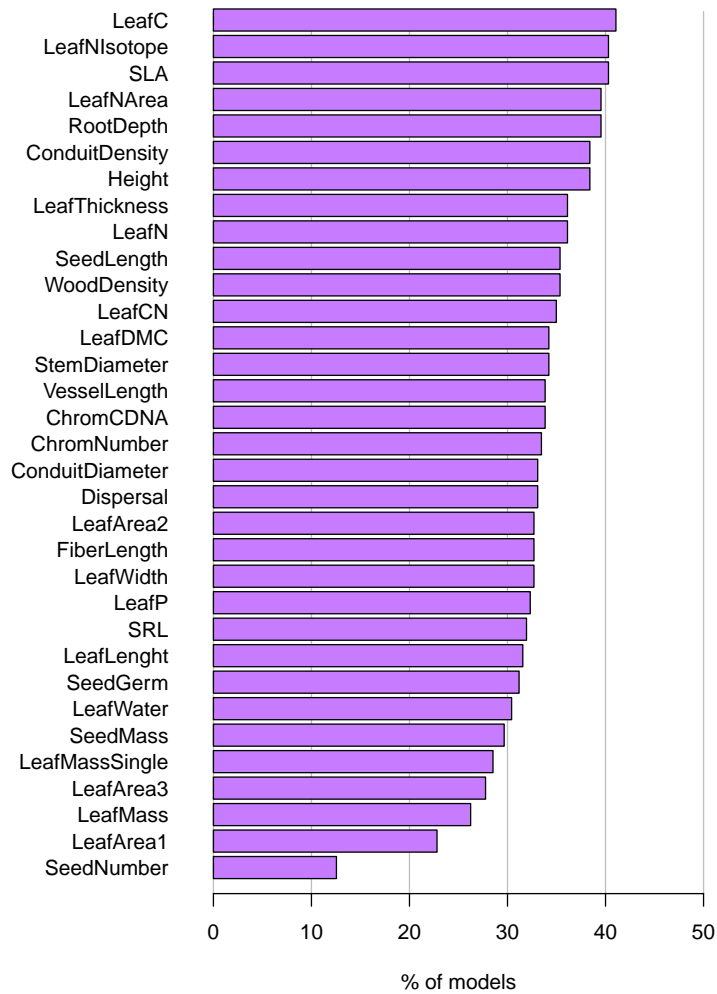
Figure 2. Ranking of different traits in cluster analyses. The ranking is based on the percent of models that include the trait in a randomized sensitivity analysis with variable number of randomly selected traits. Analyses were conducted including all traits and all F31 biome maps.
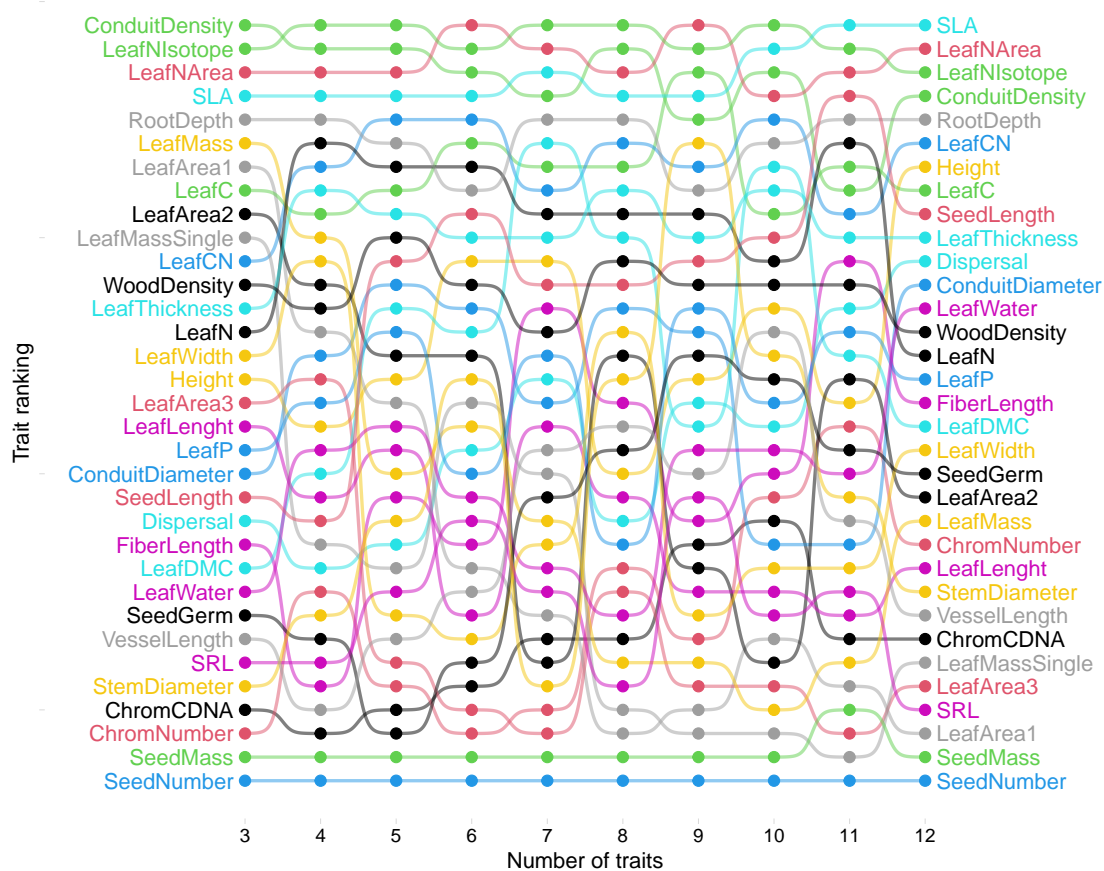
Figure 3. Ranking of traits in response to the number of traits in cluster analysis. For each number of traits, cluster analyses for all F31 biome maps were merged, and the 10% of all models with the highest κ value were selected. For these models, the number of models including each trait was counted, and counts were used to derive the ranking.
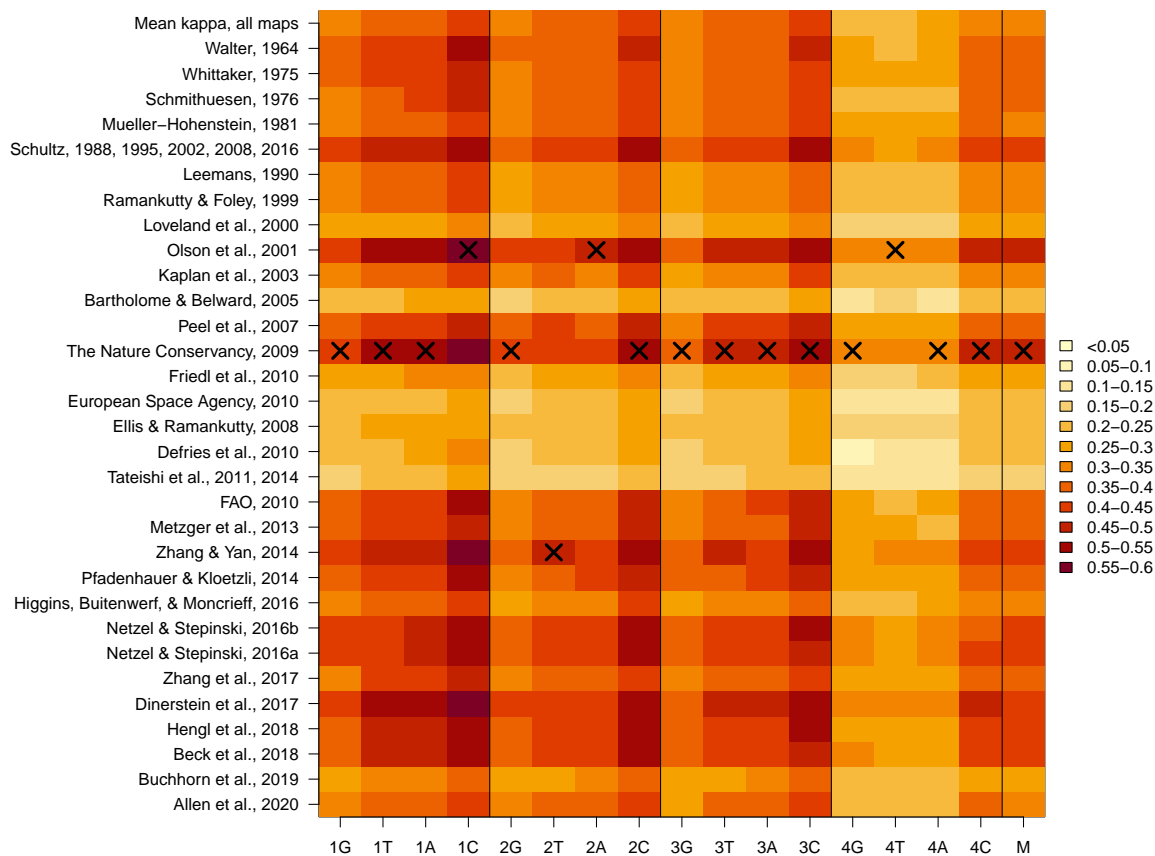
Figure 4. Data-model agreement for different biome maps and trait clusters used for the clustering. The figure shows κ values for all of the 31 biome maps provided by Fischer et al. (2022) for different trait clusters. Cluster 1 includes conduit density, leaf thickness, leaf N, SLA and rooting depth, cluster 2 includes leaf width, height, leaf N, SLA and rooting depth, cluster 3 includes height, SLA and wood density, cluster 4 includes leaf C, seed germination rate, seed number, chromosome number and wood fiber length. 'G' represents clustering with traits of grasses only, 'T' represents clustering with traits of woody plants only, 'A' represents clustering with all PFTS (grasses and woody plants), and 'C' represents clustering where both traits of grasses and woody plants (from G and T) were combined. The 'X's' denote the biome map that maximized κ within each trait cluster (each column). The first row is the mean κ value for each trait cluster averaging all biome maps, the last column is the mean κ value for each biome map, averaging all trait clusters.
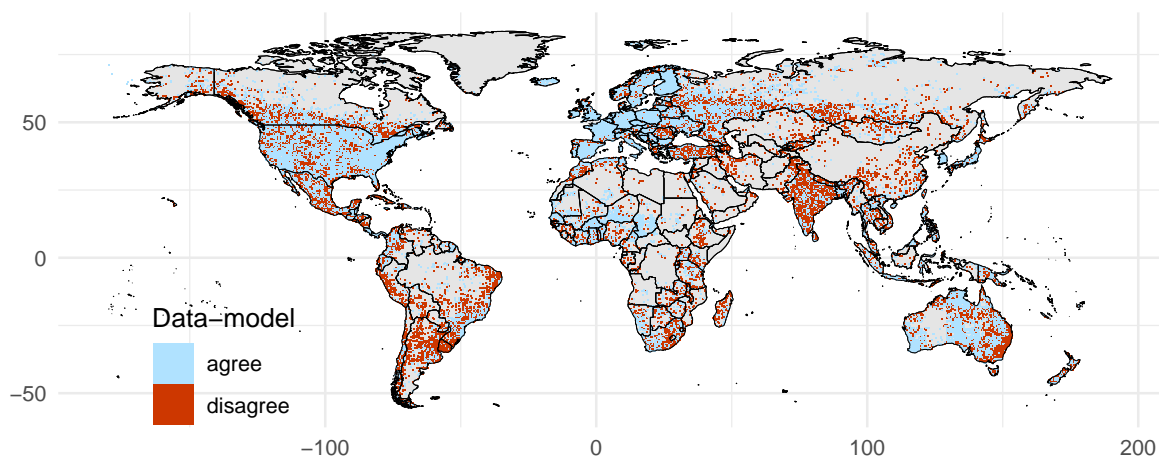
Figure 5. Agreement between an observation-based biome map and a map derived from clustering. Here, the cluster analysis with the highest κ value value was selected: cluster 1 (wood density, rooting depth, SLA, height, isotopic leaf nitrogen and conduit density) and the Nature Conservancy (2009) biome map (see Fig. 4).
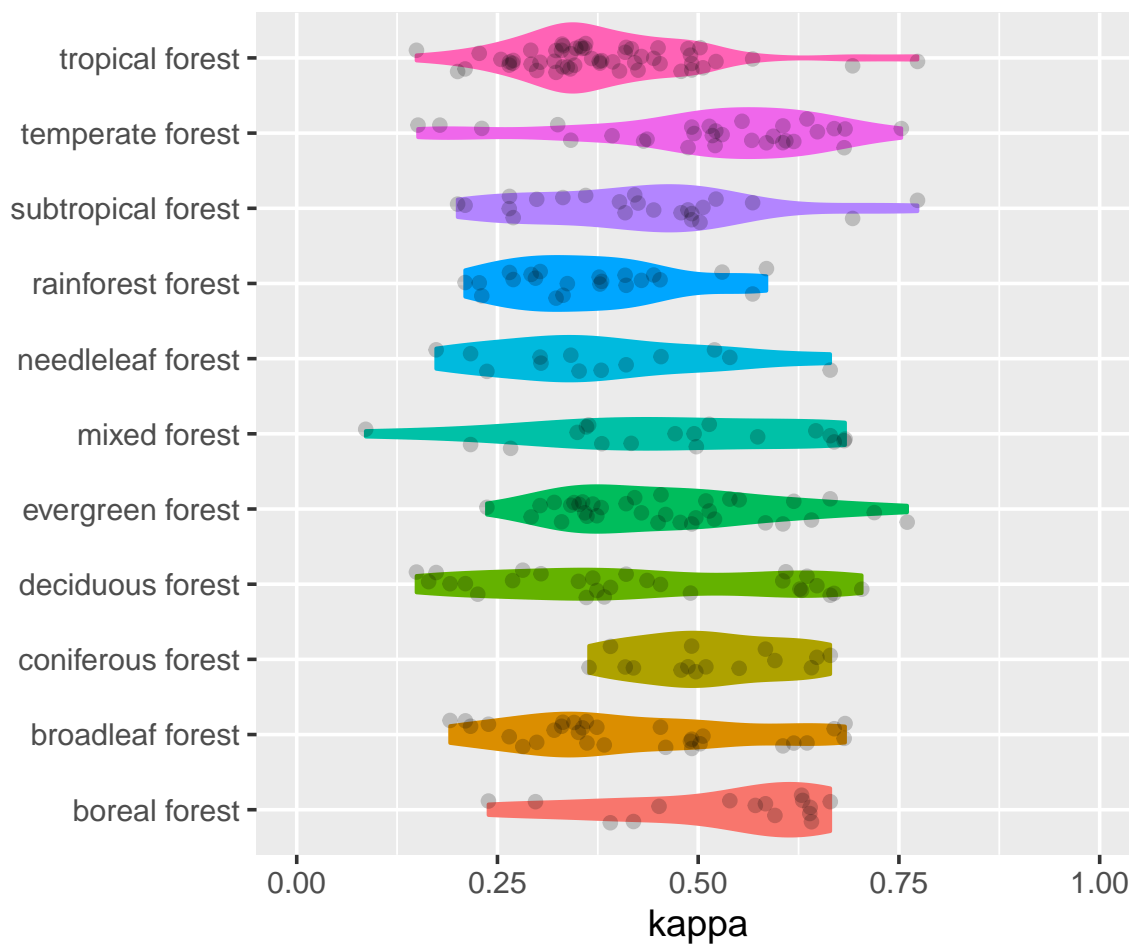
Figure 6. Model performance (κ values) in different forest types. For this analysis, all biomes that contain the attributes provided in the figure in their names were identified in the F31 biome maps. Then, the κ value was calculated for each biome (represented by the points in the figure).
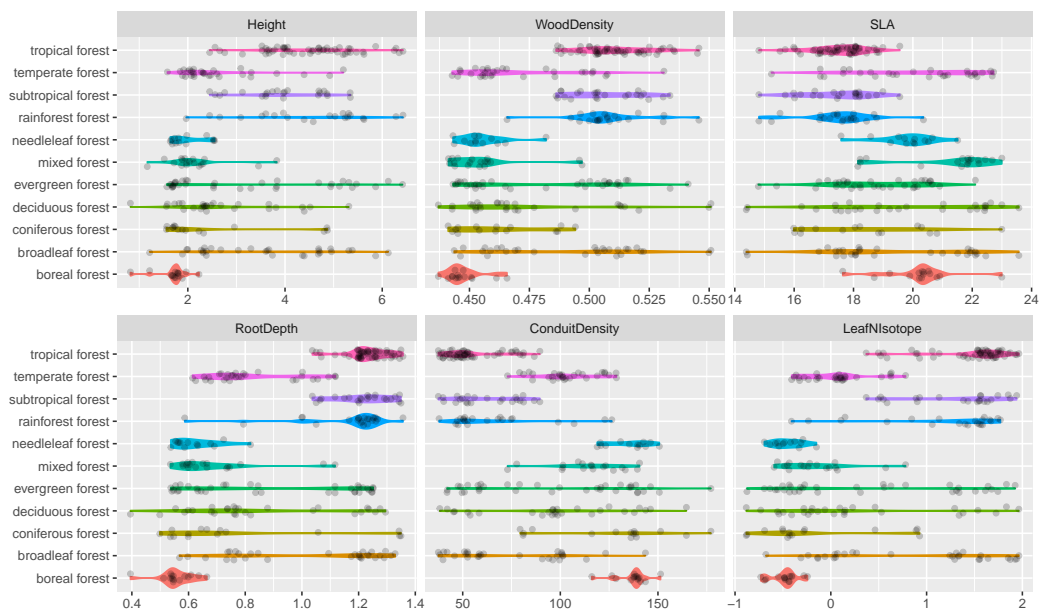
Figure 7. Mean trait values in different forest types. For this analysis, all biomes that contain the attributes provided in the figure in their names were identified in the F31 biome maps. Then, the mean trait values were calculated for each biome (represented by the points in the figure).
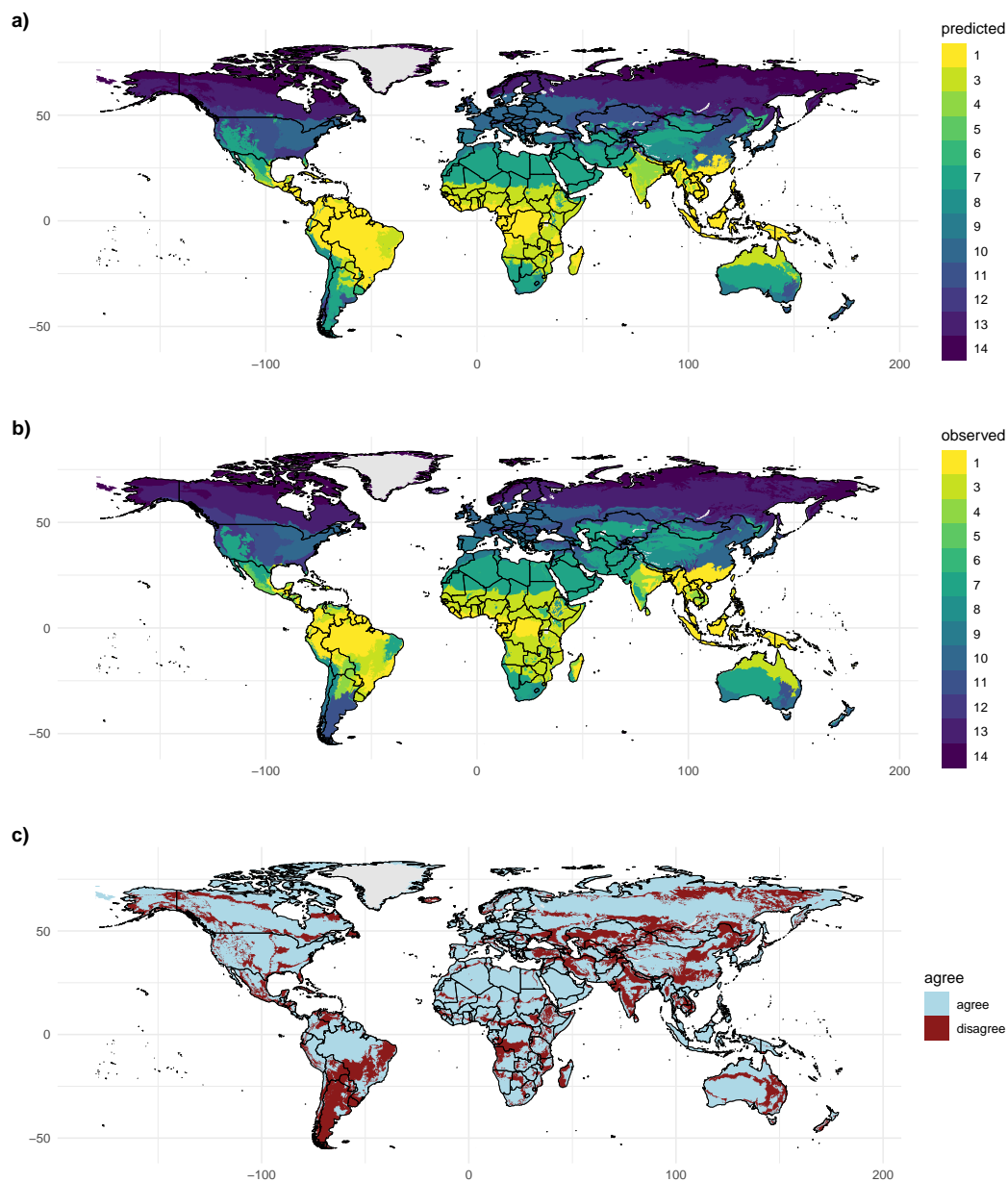
Figure 8. Global biomes derived from traits. Using species distribution models and bioclimatic data, biome patterns derived from the spatial coverage of the trait data were extrapolated to the global scale. Here, results from trait cluster 1 and the Nature Conservancy (2009) biome map were used. Biomes: 1, Tropical subtropical moist broadleaf forest; 3, Tropical subtropical grassland savanna and shrub; 4, Tropical subtropical dry broadleaf forest; 5, Tropical subtropical coniferous forest; 6, Flooded grassland and savanna; 7, Desert and xeric shrub; 8, Montane grassland and shrub; 9, Mediterranean forest woodland and scrub; 10, Temperate broadleaf and mixed forest; 11, Temperate grassland savanna and shrub; 12, Temperate conifer forest; 13, Boreal forest/taiga; 14, Tundra.