Mapping multiple dimensions of forest diversity using spaceborne spectroscopy

J. Antonio Guzmán Q.[1,2]*, Jonathan A. Knott [3], Jesús N. Pinto-Ledezma [2], Philip A. Townsend [4], Jeannine Cavender-Bares [1,2]*

[1] Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA.

[2] Department of Ecology, Evolution, and Behavior, University of Minnesota, Saint Paul, MN 55108, USA.

[3] USDA Forest Service, Northern Research Station, Saint Paul, MN 55108, USA.

[4] Department of Forest and Wildlife Ecology, University of Wisconsin—Madison, Madison, WI 53706, USA.

*Corresponding authors: J. Antonio Guzmán Q. and Jeannine Cavender-Bares

Email: aguzman@fas.harvard.edu and jcavender@fas.harvard.edu

Author Contributions: J.A.G.Q. and J.C.-B. designed research; J.A.G.Q., J.A.K., and J.N.P.-L. performed research; J.A.G.Q. contributed to analytic tools and analyzed data; J.C.-B. obtained funding; J.A.G.Q. and J.C.-B. drafted the manuscript; J.A.G.Q., J.A.K., J.N.P.-L., P.A.T., and J.C.-B. wrote the final version of the paper.

Abstract

Observing biodiversity across space and time is essential for advancing and verifying conservation efforts toward global biodiversity and sustainability goals. Spaceborne imaging spectroscopy has emerged as a revolutionary tool for quantifying and tracking forest diversity, yet its application at large spatial scales remains a central challenge. We develop a framework to map multiple dimensions of forest community composition and diversity by integrating imaging spectroscopy from two spaceborne sensors (DESIS and EMIT) with taxonomic, phylogenetic, and functional trait datasets, and 43,155 forest inventory plots across the Eastern United States. We find that spectral dissimilarity among forest communities is positively correlated with β-diversity matrices of compositional dissimilarity. We then show that imaging spectroscopy can be used to predict ordination axes of β-diversity and to map multiple dimensions of forest diversity at high spatial resolution (30 or 60 m). Predicted β-diversity axes can be used to model forest attributes, including forest types, plant lineages, and community plant traits. On average, β-diversity axes explain more than 48% of the variance—outperforming climatic and topographic predictors—and enable accurate mapping of 95 forest attributes. Our framework shows that spaceborne imaging spectroscopy, when combined with inventory data, allows indirect yet comprehensive observation of forest diversity attributes across broad spatial extents. This integrative approach sets the stage for scalable forest monitoring in support of global biodiversity conservation and forthcoming satellite missions.

Keywords: β-diversity, biodiversity, plant lineages, functional traits, imaging spectroscopy

Main Text

Introduction

Monitoring biodiversity and its changes across space and time represents a major societal challenge critical for sustainable management of our planet (1, 2). The accelerating loss of biodiversity and shifts in species composition resulting from human pressures and global change underscore the urgent need to develop effective approaches for biodiversity monitoring (3). Consequently, accurate and scalable monitoring is essential for verifying efforts towards the sustainability targets of the Convention on Biological Diversity and for assessing whether current actions are helping to “bend the curve” of biodiversity loss (4). Biodiversity observation and monitoring are particularly important in forest ecosystems, which provide essential ecosystem services and are highly vulnerable to deforestation, pests, pathogens, and abiotic stressors. Forest biomes in North America, in particular, represent a high-priority region for monitoring given their vast extent and their critical role in supporting life systems and the economy (5).

Our capacity to observe biodiversity across broad spatial scales is hindered by the inherent spatial and sampling limitations of traditional field surveys. Furthermore, field surveys are also constrained by financial and logistical challenges, including inaccessibility of remote areas or private lands (6). As a result, imaging spectroscopy has been proposed as a transformative approach for observing biodiversity and advancing its conservation (7). In particular, spaceborne spectroscopy offers a promising avenue for mapping forest diversity due to its global coverage and spatial and spectral resolutions sufficient for biodiversity assessment (8, 9). Nonetheless, many current efforts to map plant diversity have relied on airborne or near-surface platforms (10–19). Compared to these platforms, spaceborne spectroscopy typically offers coarser spatial resolution (e.g., 30–60 meters), which tends to capture forest communities rather than individual trees or crowns. In addition, airborne and near-surface approaches often cover limited regions and ecosystems due to their restricted spatial extent. Consequently, there is a clear need for spaceborne spectroscopy to fill the gap of continuous forest diversity observations over extensive areas (20), but the capability of doing so along with the multiple dimensions of forest diversity remains a central challenge.

A growing body of literature demonstrates that spatial dissimilarity in species composition among communities—known as taxonomic beta (β)-diversity—is positively associated with spectral dissimilarity

(16, 21–24). This correspondence between compositional and spectral dissimilarity has been observed using both hyperspectral airborne data and vegetation indices derived from satellite imagery. Likewise, spectral dissimilarity among species has been shown to correlate with phylogenetic and functional differences, with more distantly related or functionally distinct species exhibiting greater spectral divergence (11, 18, 19). These findings might suggest that integrating spectral dissimilarity with measures of community dissimilarity may enable the remote observation of forest diversity at broad spatial scales, particularly in regions with well-characterized ground-based data. Moreover, community dissimilarity metrics provide both a comprehensive view of forest diversity dimensions as well as insights into species turnover and nestedness by partitioning β-diversity into its replacement and richness-difference components (25–27).

Here, we present a novel framework for mapping multiple dimensions of forest community composition—taxonomic, phylogenetic, and functional—at a large spatial extent using spaceborne spectroscopy (Fig. 1). Our framework further supports the characterization and mapping of forest types, plant lineages, and community-level plant traits at broad extents through the observation of forest community composition. To achieve this, we first evaluate the relationship between multiple dimensions of β-diversity and spectral dissimilarity among forest communities using satellite data from the Eastern U.S., specifically from the DLR Earth Sensing Imaging Spectrometer (DESIS), the Earth Surface Mineral Dust Source Investigation (EMIT), and field data from the USDA Forest Service Forest Inventory and Analysis (FIA) Program. We then assess the extent to which these spaceborne observations can predict β-diversity ordinations. Our general hypothesis is that forest communities with dissimilar taxonomic, phylogenetic, or functional composition exhibit higher spectral dissimilarity than communities with similar compositions. Finally, we test the use of spectral models based on β-diversity ordinations to predict the presence or absence of forest types and plant lineages, as well as to estimate community-weighted means (CWM) of plant traits. The study provides a novel approach for mapping forest biodiversity and its multiple dimensions over large spatial extents by leveraging satellite observations from two hyperspectral spaceborne sensors and high coverage forest inventories.

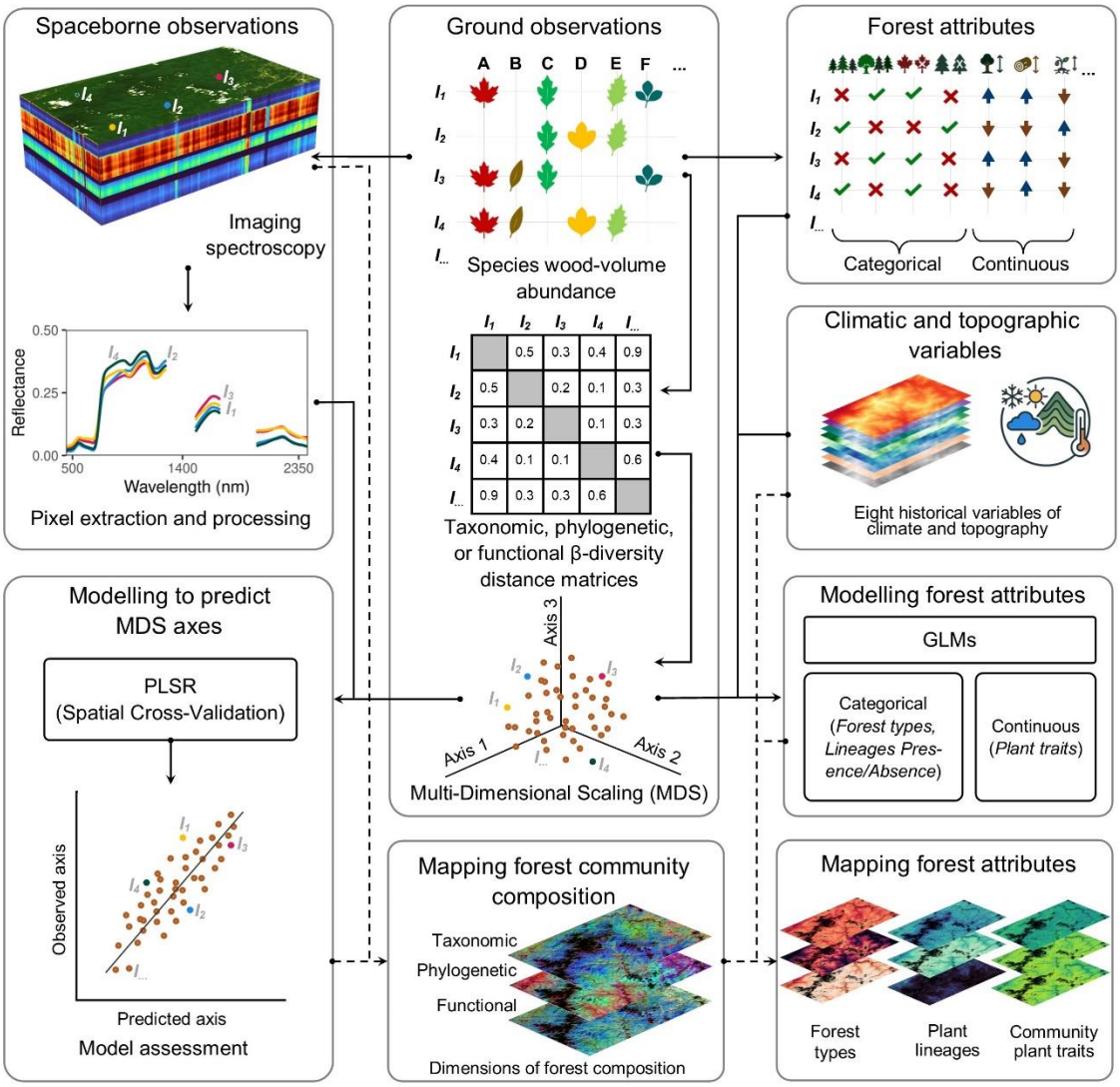


Figure 1. Framework for mapping forest community composition and its attributes using spaceborne imaging spectroscopy. Forest inventory locations are first used to extract pixels from scenes. Species abundance data from inventories are then used to calculate pairwise β-diversity matrices, which are then ordained using multi-dimensional scaling (MDS). The resulting axes of β-diversity are combined with spectral information to model and map forest community composition. These axes are also paired with climatic and topographic variables to map a range of forest attributes. Together, this framework enables spatial mapping of both forest community composition and associated forest attributes at large spatial extent.

Results and Discussion

Correspondence between β-diversity and spectral dissimilarity. We used data from 43,155 FIA plots across the Eastern United States, encompassing 15,550,151 individual trees representing 243 species, 45 genera, and 23 families (Fig. S1). For each plot, we calculated taxonomic, phylogenetic, and functional abundance-weighted β total diversity ($T\beta_{total}$, $P\beta_{total}$, and $F\beta_{total}$, respectively), and their partitioning associated with differences in species identities (i.e., $\beta_{replacement}$) or species richness (i.e., $\beta_{richness}$). From these plots, we extracted 11,526 and 10,469 clear-sky pixels from DESIS and EMIT imagery, respectively, corresponding to locations with spatial overlap with forest inventory data.

We first evaluated the relationships between spectral dissimilarity—derived using Spectral Angle Mapper (SAM)—and dimensions of β_{total}, revealing that communities differing in composition also tend to exhibit dissimilar spectral signatures (Fig. 2). These associations were generally stronger in spaceborne observations from DESIS than from EMIT, as indicated by Mantel tests. Within the DESIS dataset, correlations with spectral dissimilarity were slightly stronger for $P\beta_{total}$ and $F\beta_{total}$ than $T\beta_{total}$. Furthermore, the correspondence between spectral dissimilarity and forest composition was more pronounced when using β_{total} rather than their partitions (i.e., $\beta_{replacement}$ or $\beta_{richness}$) (Fig. S2–S3). This suggests that the combined effects of $\beta_{replacement}$ and $\beta_{richness}$—as captured by β_{total}—are more spectrally distinguishable than either component alone. Overall, the correspondence of our community-level findings are consistent with previous species-level analyses using airborne spectroscopy (11, 14, 16, 24), reinforcing the potential of using β_{total} dimensions as a framework for remotely sensed forest diversity through community dissimilarity.

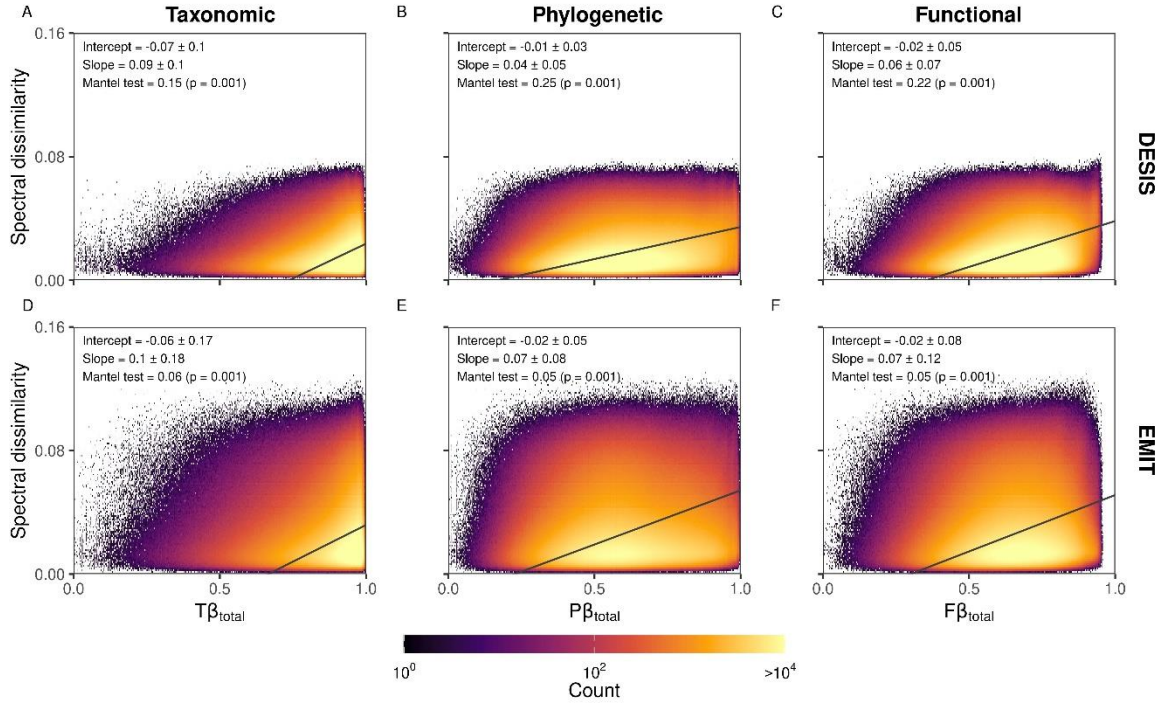


Figure 2. Relationship between spectral dissimilarity and dimensions of beta total diversity (βtotal) in relation to taxonomic ($T\beta_{total}$), phylogenetic ($P\beta_{total}$), and functional ($F\beta_{total}$) information. The black solid lines represent the average linear regression line when comparing dissimilarities from a target community with all communities. Extended figures are shown in Fig. S2–S3.

β-diversity patterns across the Eastern United States. To address the high dimensionality of the β_{total} matrices for modeling purposes, we applied Multi-Dimensional Scaling (MDS) as a data reduction technique, reducing β_{total} matrices into three axes. This approach revealed patterns of dissimilarity in forest community composition and their associated spatial distributions (Fig. 3). Although all ordinations were derived from the same inventory data, MDS axes based on $P\beta_{total}$ and $F\beta_{total}$ information exhibited stronger goodness-of-fit (R^2 = 0.71 and 0.60, respectively) than MDS axes based on $T\beta_{total}$ (R^2 = 0.11), when correlating pairwise ordination distances with the original β_{total} matrices (Fig. S4). These results suggest that the structure of forest community composition is more effectively captured when incorporating phylogenetic and functional dimensions, rather than relying solely on taxonomic identities. Nonetheless, MDS axes were highly correlated across diversity dimensions within the same ordination space (Fig. S5), indicating that the relative positioning of communities is largely consistent among the different dimensions of diversity. This convergence also implies that β_{total} ordinations axes are capturing congruent patterns of community differentiation, where evolutionary history and ecological function are closely linked and likely shaped by shared underlying processes (27, 28).

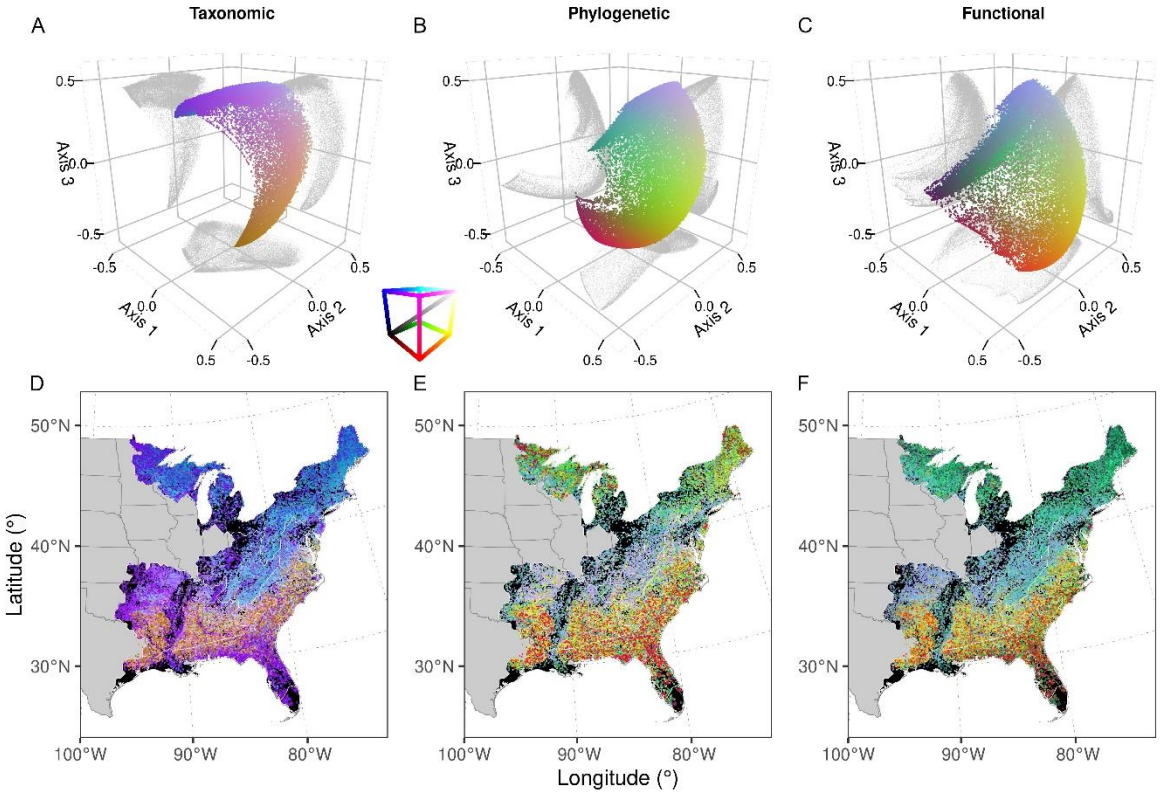


Figure 3. Multi-Dimensional Scaling (MDS) to ordinate forest communities according to their taxonomic (A and D), phylogenetic (B and E), and functional (C and F) β-diversity and their spatial variation across the Eastern U.S. Gray shadows (A-C) represent the projection of three-dimensional axes of β-diversity onto two-dimensional planes.

The projection of weighted average scores and the correlation of climatic and topographic variables with the MDS ordinations (Fig. S6–S9) revealed the presence of underlying climatic and topographic gradients that likely influence forest community compositional dissimilarity. Our analyses suggested that across all three ordinations, higher mean annual temperature (MAT) and lower values of temperature annual range (TAR) were associated with the higher values of the first MDS axes. In contrast, lower precipitation seasonality (PS) and higher mean annual precipitation (MAP) were associated with the second axes, while elevation and slope correlated with the third axes. The positive associations of the first two MDS axes with MAT and MAP are consistent with expectations based on the MDS rotation

procedure (see Materials and Methods for details). Furthermore, the MDS ordinations appeared to capture gradients in the CWM of plant traits and the relative abundances of plant lineages, as indicated by the projection of weighted average scores and correlations with MDS axes (Fig. S6 –S9). Collectively, these findings support the interpretation that the MDS ordinations of β_{total} effectively capture dimensions of forest community composition across environmental gradients.

Predicting ordinations of forest community composition from spectra. We developed a workflow (Fig. S10) to predict MDS ordination axes for each dimension of β_{total} using observations from DESIS and EMIT and by applying repeated Partial Least Squares Regression (PLSR) within a machine learning framework. Our results demonstrate that it is feasible to predict MDS β-diversity axes across all ordinations (Fig. S11 – S14), highlighting the potential to map multiple dimensions of forest community composition at large spatial scales. Model performance varied considerably across diversity dimensions, MDS β-diversity axes, and sensors, with training and testing R^2 values ranging from 0.08 to 0.64 and percent RMSE (%RMSE) between 17% and 29% (Tables S2 and S3). Across all models, Axes 1 consistently yielded higher R^2 than Axes 2 but not Axes 3. Predictions for Axes 1 required more latent components to achieve optimal predictions than Axes 2 or 3 (Fig. S15). Between sensors, models based on EMIT data often outperformed those based on DESIS (Table S2 and S2).

Despite the observed differences in model performance, the Variable Importance in Projection (VIP) revealed consistent spectral predictors across sensors, diversity dimensions, and MDS axes (Fig. 4). The similarity in VIP scores across both DESIS and EMIT sensors, as well as among the different MDS axes, underscores the robustness and generalizability of spectral-diversity relationships captured by our models. This convergence indicates that spectral regions—particularly those around 710, 790, and 1521 nm—consistently contribute to the prediction of community composition across taxonomic, phylogenetic, and functional dimensions. These wavelengths are generally linked with vegetation greenness, forest structure, water content, respectively suggesting potential spectral features that underpin key biophysical and biochemical aspects of community composition (Table S4). The alignment of VIP profiles among sensors suggests that forest composition signals are not sensor-specific, but instead reflect key spectral regions that are broadly detectable from space. This finding highlights the potential for transferring models across sensors and supports the development of a unified, sensor-agnostic framework for large-scale diversity monitoring. Furthermore, the consistency of VIP scores across MDS β-diversity axes points to ecologically meaningful gradients that are spectrally sensed, reinforcing the utility of spectroscopy information as a proxy for mapping forest composition. Collectively, these patterns strengthen the feasibility of operationalizing forest biodiversity monitoring using current and future spaceborne spectroscopy missions such as NASA's SBG (Surface Biology and Geology) and ESA's CHIME (Copernicus Hyperspectral Imaging Mission for the Environment).

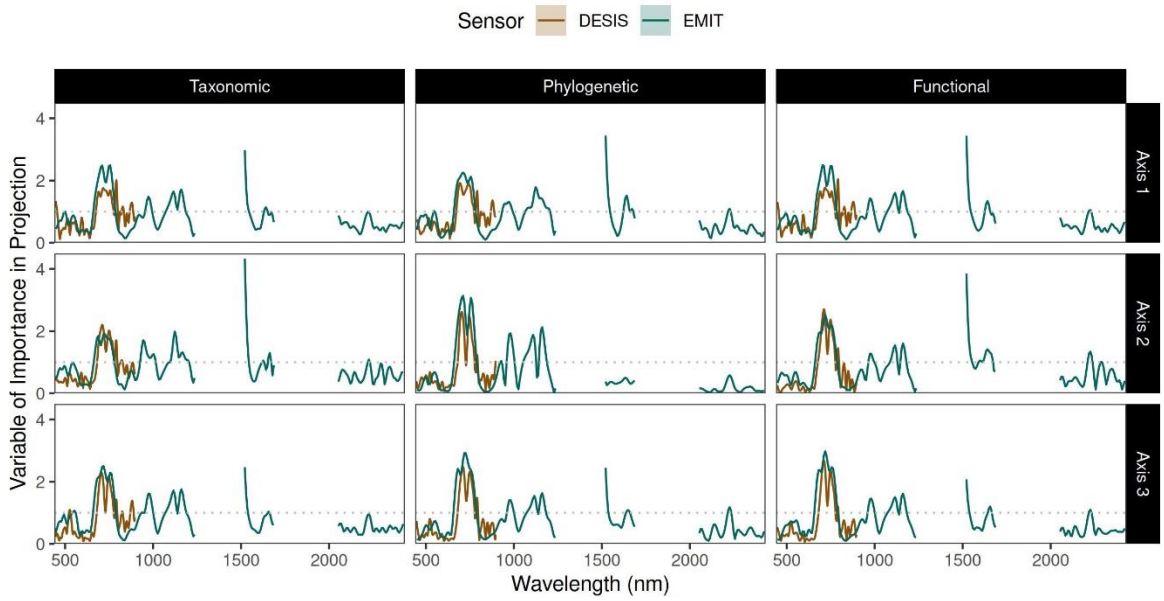


Figure 4. Variable Importance of Projection across wavelengths of PLSR models to predict MDS axes of β-diversity at different dimensions using spaceborne observations of DESIS and EMIT.

Predicting forest attributes. We integrated MDS axes with climatic and topographic variables using generalized linear models (GLMs), revealing that distinct forest attributes—including forest type, the presence or absence of plant lineages, and CWM of plant traits—can be effectively modeled and predicted (Fig. S16–S18). The average True Skill Statistic (TSS) for binomial models predicting presence/absence in testing datasets was 0.6868 ± 0.144 (Area Under the Curve [AUC] = 0.900 ± 0.066) across 59 forest types, while for 28 plant lineages the performance was slightly lower (TSS = 0.54 ± 0.19, AUC = 0.84 ± 0.10). Similarly, the average R^2 for continuous models predicting CWM of eight plant traits was 0.46 ± 0.20 (%RMSE = 15.18 ± 2.03). The relative importance of variables in these GLMs indicated that three MDS axes derived from β_{total} dimensions accounted for more than 48% (±13.11%) of the total variable importance across forest attributes. In contrast, four climatic and three topographic variables contributed 36.48% (±11.35%) and 15.11% (±7.96%), respectively (Fig. S19–S21). The greater explanatory power of MDS β-diversity axes over climatic or topographic variables underscores the importance of community context for predicting forest attributes. By capturing β-diversity dimensions, MDS ordinations are likely to reflect not just environmental conditions but also the outcomes of species interactions and ecological filtering embedded within species distributions. These findings indicate that predictions of forest attributes are shaped more by community composition than by abiotic constraints alone, interpreting spectral signals based on realized species distributions and forest composition, rather than fundamental niches (29, 30). As a result, MDS axes of β-diversity provide an ecologically grounded basis for modeling and mapping forest attributes.

By applying GLM coefficients to MDS β-diversity axes predicted from spaceborne spectroscopy at the overlapping inventory locations, our workflow demonstrated that a wide array of forest attributes can be inferred from space (Fig. 5 and Fig. S22–S24). The accuracy of these predictions was evaluated by regressing expected probabilities (for categorical attributes) or CWMs (for continuous traits) from the GLM models against their predicted values. This validation approach revealed an average R^2 of 0.69 ± 0.23 (%RMSE = 16.98 ± 5.51) for forest types, 0.70 ± 0.22 (%RMSE = 15.17 ± 7.58) for plant lineages, and 0.49 ± 0.24 (%RMSE = 16.83 ± 3.16) for community plant traits across testing datasets, independent of sensor platform. Although sensor-specific performance varied for some forest attributes, overall model accuracy was relatively consistent between sensors. In general, these results highlight the potential of combining spectral data and β_{total} ordinations to generate ecologically meaningful, spatially continuous

predictions of forest community composition—paving the way for scalable biodiversity assessments from space.

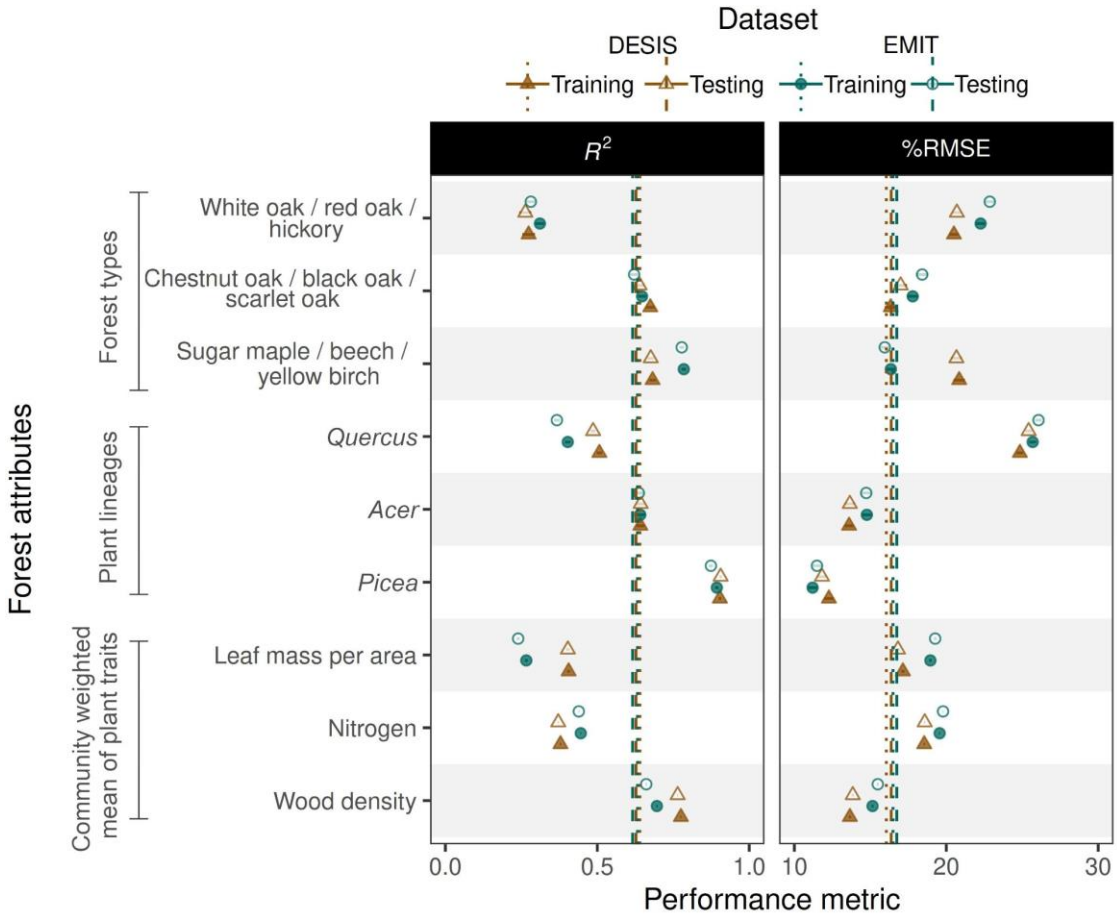


Figure 5. Model performance for predicting attributes of forest composition from spaceborne observations from DESIS or EMIT. Each point represents an average of 100 repeated models. Horizontal lines represent the average performance for all forest attributes. Extended figures are shown in Fig. S22–S24.

Mapping forest community composition and their attributes. We used hyperspectral scenes of DESIS and EMIT and applied model coefficients from the PLSR and GLM analyses to predict and map dimensions of forest composition across the eastern U.S. and to estimate a suite of associated forest attributes at moderately high spatial resolution (30 m or 60 m). Here, we illustrate our workflow using an EMIT L2A scene from the southern Appalachian Mountains, spanning parts of North Carolina and Tennessee, USA (Fig. 6). In this region, variation in forest community composition—captured by predicted MDS β-diversity axes—aligns closely with elevation gradients. Map values and color scales correspond to those in Fig. 3, enabling the back-projection of mapped communities within the ordination space. We computed the uncertainty of prediction for each pixel in these composition maps as the summed amplitude of predictions. In this particular scene, the uncertainty in the composition tends to co-vary across dimensions, where higher values are observed in lowland areas likely due to spectral mixing of frequent non-forest covers. Leveraging the maps of community composition in combination with climatic and topographic variables, we then generated spatial predictions of various forest attributes. For example, the probability of occurrence of specific forest types was modeled using taxonomic composition maps, while maps of phylogenetic composition informed predictions of plant lineage occurrence. Similarly, functional composition maps enabled spatial estimation of CWM of plant traits. In this region, the high abundance of broadleaf deciduous communities at lower elevations (i.e., below ~1,000 m a.s.l.)

corresponds with the high predicted occurrence of *Quercus* (oak) species and oak-dominated forest types in the scene. These areas are also characterized by elevated foliar nitrogen (N) concentrations and lower predicted values of leaf mass per area (LMA). In contrast, at higher elevations—primarily above 1,800 m a.s.l.—there is a greater prevalence of evergreen coniferous communities, marked by high predicted occurrence of *Picea* (spruce) species, high LMA values, and lower foliar N concentrations. Mid-elevational zones exhibit a transitional pattern, with forest communities dominated by *Acer saccharum* (sugar maple), *Betula alleghaniensis* (yellow birch), and *Fagus grandifolia* (American beech) to mention some. Ultimately, these maps offer a scalable solution for capturing forest diversity and composition across multiple ecological dimensions and broad spatial scales, with significant potential to enhance our capacity to monitor and understand forest biodiversity in support of sustainability and stewardship goals.

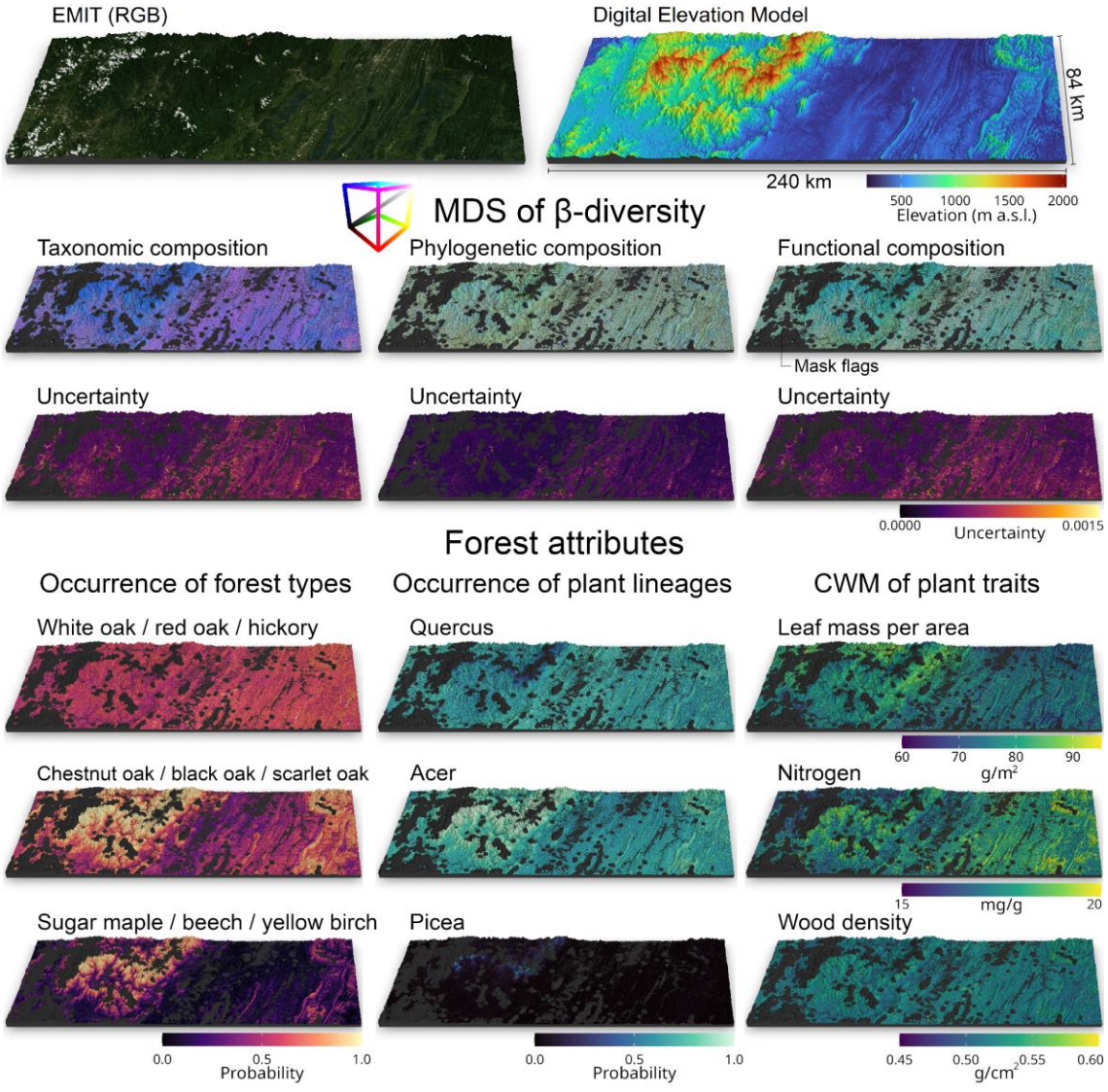


Figure 6. Mapping of multiple dimensions of forest diversity using EMIT across the southeast of the United States. Top panels represent a true color image (R: λ_{663}, G: λ_{551}, B: λ_{425}) and the elevation model provided by NASA with EMIT imagery. Middle panels of community composition represent the mapping of MDS β-diversity axes based on different dimensions of diversity as well as their uncertainty. Bottom panels describe the mapping of forest attributes associated with the occurrence of forest types, plant

lineages, and CWM of plant traits using the predicted MDS β-diversity axes, climatic, and topographic variables. Each pixel value represents the average of 100 repeated models.

Conclusions

Our study presents a comprehensive methodological framework for mapping dimensions of forest diversity by integrating thousands of forest inventory plots with spaceborne imaging spectroscopy. Spaceborne spectroscopy offers considerable potential for enhancing our ability to observe and monitor forest biodiversity. Yet mapping forest diversity across large spatial extents—given diverse ecosystem patterns, high species richness, and major shifts in species composition—presents significant challenges. Our findings demonstrate that across the Eastern U.S., forest communities with dissimilar compositions (e.g., high β-diversity) exhibit corresponding spectral dissimilarities. The correspondence between spectral dissimilarity and compositional dissimilarity is consistent across multiple dimensions of β-diversity (i.e., taxonomic, phylogenetic, and functional), for both components of β-diversity partitioning (i.e., replacement and richness), and for spectral information from two spaceborne sensors (i.e., DESIS and EMIT), providing a solid foundation for using in situ β-diversity to remotely sense forest diversity from space. Due to their complexity and high dimensionality, however, pairwise β-diversity matrices are challenging to apply directly to modeling biodiversity from satellite spectra. To address this challenge, our framework employs ordinations of β-diversity, in which the combination of three axes captures spatial patterns of forest community composition. We show that ordination axes of β-diversity can readily be used to predict biodiversity from imaging spectroscopy, and that their application to satellite observations allows the spatial mapping of multiple dimensions of forest community composition from space. Moreover, our modeling approach reveals regions of the electromagnetic spectrum that are consistently important for predicting axes of β-diversity across multiple diversity dimensions and for both spaceborne sensors, underscore the potential for models based on imaging spectroscopy to capture signals of forest composition and to generate a sensor-agnostic framework for large-scale diversity mapping. By integrating a range of forest attributes with axes of β-diversity across different diversity dimensions, along with climate and topographic variables, we further demonstrate that our framework can effectively map forest types, plant lineages, and community plant traits. This approach reveals that forest community composition—captured through β-diversity axes—plays a stronger role than climate or topography, thus offering a more ecologically grounded basis for mapping species occurrence and functional traits. The effectiveness of our framework for mapping forest community composition and diverse forest attributes at high spatial resolution and across large spatial extents sets the stage for the verification of forest management efforts aimed at sustaining planetary biodiversity and resilience.

Materials and Methods

Study area. Our study was conducted in the Eastern continental U.S. region with an area close to 2,593,107 km^2 (Fig. S1). This region encompasses eight eco-climatic domains with distinct vegetation, landforms, and ecosystem dynamics defined by the National Ecological Observatory Network (NEON) (31). Despite its extension, this region presents a comparable timing in the peak of greenness (i.e., June through August) (32, 33), which can be used to restrict the phenological effects on the optical properties of forest communities.

Forest inventory data. We used inventory data collected from the Forest Inventory and Analysis (FIA) program of the United States Department of Agriculture Forest Service (34). We used 43,155 forested plots in the region of interest collected between January 1st, 2018, and January 12th, 2023. Each inventory plot consists of a central circular subplot (7.31 m radius) surrounded by three circular subplots at 36.57 m from their centroids located at 0-, 120-, and 240-degree azimuths (35). Trees (> 12.7 cm diameter at breast height) from all the subplots were used to characterize local tree communities given the spatial resolution of the spaceborne observations. We estimate wood volume for each live tree assuming a cylinder shape using the diameter at breast height and tree height, and then expand its value to per-unit-area using the FIA trees-per-acre expansion factor. The resulting wood volume per unit of area was then summed per species as a descriptor of species abundance for further analysis. We used the forest type

with the largest proportion of occurrence that is recorded on each FIA plot as a descriptor of the forest community for further analysis. More details about the selection of inventories and filtering of plots are shown in Methods S1.

Taxonomic, phylogenetic, and functional β-diversity. We estimated β total diversity (β_{total}) following improvements by (26) based on pioneering work by (25) as a framework to estimate the taxonomic, phylogenetic, and functional dissimilarities among tree communities. Within this framework, pairwise comparisons of communities were performed to partition β-diversity into components of replacement ($\beta_{replacement}$, i.e., diversity explained by replacement of species alone) and richness ($\beta_{richness}$, i.e., diversity explained by species loss/gain alone). These were used to compute β total diversity (β_{total}) as the sum of $\beta_{replacement}$ and $\beta_{richness}$ (See more details in (26)). We used Jaccard dissimilarity weighted by the relative abundance of the wood volume per unit of area in all the estimations of β. Specifically, we used a pruned phylogenetic tree from (36) to estimate phylogenetic β (Pβ) (Fig. S25). The phylogenetic tree was obtained through the 'phylo.maker' function of the V.PhyloMaker2 package (37) in R (38) using its third scenario. Under the third scenario, V.PhyloMaker2 adds missing species to the half point of their parent branch and uses the BLADJ approach for branch length estimation. For functional β diversity (Fβ), we used eight plant traits for each species (Methods S2 and Fig. S25) to summarize their variation into three principal components (Fig. S26) and construct a dendrogram forced as a phylogeny. Taxonomic β diversity (Tβ) was estimated using species taxonomic names occurring in the plots. β values were estimated using a modification of the BAT package by (39) with C^{++} and OpenMP as a backend to efficiently run a large number of pairwise comparisons between communities in parallel using high-performance computing.

Ordinations of β-diversity. To reduce the dimensionality of the taxonomic, phylogenetic and functional trait matrices created from the forest inventory data, we applied landmark Multi-Dimensional Scaling (MDS) to ordinate communities in three-dimensions (i.e., axes), according to their β_{total} for each dimension of forest diversity. Each MDS was rotated to match mean annual temperature and annual precipitation gradients derived from Worldclim 2 (40), using a sequential orthogonal rotation procedure that aligns the ordination axes with the fitted environmental vectors via vector fitting and planar rotations. To evaluate each ordination and their meaning, we first calculated the goodness of fit as the proportion of variance (R^2) of β_{total} that is accounted for by euclidean distances in MDS β-diversity axes values between communities. Then, climatic, topographic, community weighted mean (CWM) of plant traits, and the relative abundance of major plant lineages (i.e., gymnosperms, angiosperms, arbuscular mycorrhizal-, ectomycorrhizal-symbiosis trees) were projected into these ordinations by computing the weighted means based on the axes using the 'wascores' function of *vegan* (41). We used MDS β-diversity axes to model dimensions of forest community composition from spaceborne observations, and to estimate then community plant traits and the occurrence probability forest types and plant lineages.

Spaceborne hyperspectral data and processing. We used scenes from two imaging spectrometers that are docked to the International Space Station: the DLR Earth Sensing Imaging Spectrometer (DESIS) (i.e., 30 m pixel resolution) and the Earth Surface Mineral Dust Source Investigation (EMIT) (i.e., 60 m pixel resolution). For both sensors we used Level 2A hyperspectral scenes with less than 70% of cloud cover that provide surface reflectance data corrected by different means (e.g., (42, 43)). For DESIS, we only employed scenes (n = 3145) within the peak of the growing season between June 15th to August 15th from 2019 to 2023. For EMIT, on the other hand, we only used scenes (n = 349) collected between June 1st to August 30th of 2023 and 2024. The spectral reflectance from both sensors were transformed using continuous wavelet transformation (CWT) in order to enhance absorption features, reduce angular/illumination effects, and smooth the spectral signal (44). For this, we first resample the reflectance spectra to a continuum band spacing (DESIS: 3.00 nm; EMIT: 7.43 nm) using the Full-Width-Half-Maximum method. CWT was applied then on the resampled spectra by selecting different scales (DESIS: 2^3, 2^4, and 2^5; EMIT: 2^2 and 2^3) that capture the combination of small and large reflectance features. These scales were summed to create a summed-wavelet spectra. Bands from the transformed spectra with potential noise, atmospheric contamination or close to the edge of the spectral range of the sensors were removed. We ended with scenes of 173 bands with a spectral range between 449 – 965 nm

for DESIS, while scenes of 179 bands with spectral ranges between 448 – 1236, 1521 – 1686, and 2053 – 2421 nm for EMIT. Our analyses excluded pixels with Normalized Difference Vegetation Index (NDVI, $\sim\lambda_{800}$ - $\sim\lambda_{680}$ / $\sim\lambda_{800}$ + $\sim\lambda_{680}$) and NIR values ($\sim\lambda_{800}$ nm) lower than 0.4 and 0.3, respectively, to ensure that observations come from vegetated surfaces and exclude shadows, respectively.

Relationship between β-diversity and spectral dissimilarity. We estimate spectral dissimilarity between plots in the inventory data using the Spectral Angle Mapper (SAM) method (45). Using SAM, reflectance spectra are n-dimensional vectors (i.e., number of bands) to estimate the spectral angle (i.e., 0° – 90°) between two communities and then estimate a metric of dissimilarity (i.e., 0 – 1), where values close to 1 describe forest communities with contrasting spectra. Using each inventory as a reference, we computed pairwise matrices of spectral dissimilarity between all plots (Methods S3). We performed Mantel tests to assess the association between matrices of spectral dissimilarity and dimensions of β and their partitions using Pearson correlations and 999 permutations. Then, using each community as a reference, we fitted separate linear models of spectral dissimilarity against multiple dimensions of β and their partition (i.e., β_{total}, $\beta_{replacement}$, $\beta_{richness}$), based on pairwise comparisons with all other communities.

Modeling multiple dimensions of β-diversity. We modeled multiple dimensions of β-diversity using FIA plots that overlap with the available transformed hyperspectral scenes. For this, we extracted pixels to predict MDS scores for each dimension of β_{total} using a modeling framework based on partial-least squares regression (PLSR). The extraction of pixels was done using the federally protected locations of FIA plots obtained from the National Information Management System internal to the FIA program. FIA plots may have multiple recorded plot locations from different remeasurement years (with GPS errors up to 10 m, (46)), so to link to the imagery, we calculated the mean latitude and longitude from all remeasurements (when available) of each individual FIA plot following (47). From DESIS scenes, we restrict the pixel extraction of overlapping inventories with less than ± 6 years of difference between the inventory and observation date. For each plot, we limit the number of extracted scene observations by selecting the clear-sky observation closest to the inventory date. For further analysis we employed a total of 11 526 and 10 469 pixels from DESIS and EMIT scenes, respectively (Fig. S1).

Our modeling framework consisted of repeated (n = 100) PLSRs using a spatial cross-validation strategy based on stratified random sampling. For this, we first randomly split 60% of the samples available per county for algorithm training. Then we applied spatial cross-validation models using a subset of training samples on each iteration. For each iteration and axis, we selected 85% of training samples, sampling randomly across histogram distributions to capture the range of axis variation along with nominal breaks. We used a 10-fold leave-locations-out cross-validation where a series of samples from counties are spatially excluded to train the model. This repeated framework helps us to compute the mean and SD of the model estimates, and thus the potential uncertainty of the predictions. Once the repeated models were computed, we determined the optimal number of components based on the spatial cross-validation error of the RMSE of prediction (48, 49). With the optimal number of components, we then estimate the model performance on both training and testing datasets by computing the coefficient of determination (R^2), the bias, the root mean squared error (RMSE), and the percentage RMSE (%RMSE = RMSE/ range of 0.99 and 0.01 quantiles × 100). From these repeated models, we extracted variable importance in projection (VIP) scores to assess which spectral regions were most important for predicting MDS β-diversity axes, and PLSR coefficients to apply to spaceborne imagery for mapping forest community composition (Fig. S27).

Modeling forest attributes. We modeled the occurrence of forest types and plant lineages, and the variation community weighted mean (CWM) of plant traits using MDS β-diversity axes, as well as climatic and topographic variables (Methods S4). To generate these models, we followed an approach similar to (50), but we focused on occurrence probabilities and CWM of plant traits. Specifically, we used $T\beta_{total}$-base MDS ordinations to predict true presence or absence of forest types, $P\beta_{total}$-base MDS ordinations to predict the true presence or absence of a plant lineage (e.g., plant family, plant genus, or mycorrhizal symbiosis) within a community, and $F\beta_{total}$-base MDS ordinations to predict the CWM of plant traits. Forest types and plant lineages were treated as dependent binary variables, while the CWM of each plant

trait was treated as a dependent continuous (Gaussian) variable for the models. For this, we trained repeated generalized linear models (GLMs) following a similar machine learning framework for modeling β-diversity. Overall, we first randomly split 60% of all forest inventories available per county for training purposes. Then, we train GLMs (n = 100) using 10-fold cross-validation for each forest type, plant lineage or plant trait using 85% of the training samples on each iteration. To overcome the imbalance of observations of scarce forest types and plant lineages, we applied the random over-sampling examples technique (51) on each iteration to aid the binary classification. Once the models were trained, we assessed their performance using I) training and II) testing datasets, and using III) predicted MDS axes from spaceborne spectroscopy on datasets that were used to train previous PLSR models, and IV) predicted MDS axes from spaceborne spectroscopy on datasets that were used validate previous PLSR models. For I and II, categorical models of forest type and plant lineages were evaluated using true skill statistic (TSS), sensitivity, specificity, and the area under the curve (AUC) of the receiver operating characteristic curve (ROC), while continuous models of CWM of plant traits were evaluated using R^2, bias, RMSE, and %RMSE. For III and IV, all models were evaluated using R^2, bias, RMSE, and %RMSE through the observed-predicted occurrence probability of forest type or plant lineages as well as the observed-predicted CWM of plant traits.

Mapping β-diversity and forest attributes. We applied our predictive models to map dimensions of forest community composition using scenes of DESIS and EMIT (i.e., Level 3 products), and then map forest attributes associated with the occurrence probability of forest types, lineages and CWM of plant traits (i.e., Level 4 products). We first applied the repeated PLSR models on DESIS and EMIT scenes to generate maps of the mean estimated MDS axes associated with dimensions of community composition. The uncertainty of these estimates per MDS β-diversity axis and dimension of diversity were mapped as the amplitude between the upper and lower limits of the confidence intervals at 95% of the predicted values following (52), as well as the sum of amplitudes among all axes from each dimension as a descriptor of the overall uncertainty. We then applied the coefficients from the repeated GLMs on these MDS maps in addition to climatic and topographic layers to generate maps of forest attributes and their uncertainties as described above.

We showcase an EMIT scene across the southern Appalachian Mountains in Tennessee and North Carolina to illustrate the potential of our workflow to map multiple dimensions of forest diversity. This scene encompasses sections of one of the most biodiverse areas of North America, including the Great Smoky Mountains National Park, Nantahala National Forest, and Frozen Head State Park. We highlight probability predictions of three of the most abundant forest types and plant lineages according to FIA plots in the region as well as predictions of CWM of leaf mass per area, wood density, and leaf nitrogen concentration. The application of our mapping efforts to all the scenes used from DESIS and EMIT are available at Harvard Dataverse.

Data, Materials, and Software Availability

Forest inventory data were obtained from the USDA Forest Service's FIA Program and are available through the FIA DataMart (https://apps.fs.usda.gov/fia/datamart/datamart.html). However, as noted in the Methods, we used a federally protected version of the FIA database to access actual plot locations for our analyses (for more information on federally protected FIA data, see https://research.fs.usda.gov/programs/fia/sds). All code associated with this research is available on GitHub (https://github.com/Antguz/mapping-communities), and will be archived in Zenodo under version 1.0 upon publication. Data that do not compromise federally protected information are being prepared for release in the Harvard Dataverse. The spaceborne data products developed in this research are also available through the Harvard Dataverse (https://dataverse.harvard.edu/previewurl.xhtml?token=cfb44b92-ec7f-4cd8-9c7c-c2b4b43612d6).

Acknowledgments

We thank the German Aerospace Center (DLR) and the National Aeronautics and Space Administration (NASA) for providing access to DESIS and EMIT imagery for scientific research, respectively. We thank the numerous field crews and information management staff who collected and compiled the FIA data used in this study. This study was supported by the NASA Biodiversity Program (Award number: 80NSSC21K1349) and the National Science Foundation (NSF) through ASCEND Biology Integration Institute (DBI: 2021898). J.N.P.-L. and P.A.T were supported by the NSF Division of Environmental Biology (DEB: 2017843) and NASA's EMIT Science program (Award number: 80NSSC24K0777), respectively. The main author thanks Vinicius Marcilio-Silva and Steven Augustine for the insightful discussions on imputing plant traits and the ecology of the manuscript, respectively. Thanks to the Minnesota Supercomputing Institute and the Faculty of Science Research Computing for their high-performance infrastructure that allows us to conduct this research. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and should not be construed to represent any official USDA or U.S. Government determination or policy.

References

1. J. Cavender-Bares, *et al.*, Integrating remote sensing with ecology and evolution to advance biodiversity conservation. *Nat Ecol Evol* **6**, 506–519 (2022).
2. A. Gonzalez, *et al.*, A global biodiversity observing system to unite monitoring and guide action. *Nat Ecol Evol* **7**, 1947–1952 (2023).
3. IPBES, “Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services” (Zenodo, 2019).
4. D. Leclère, *et al.*, Bending the curve of terrestrial biodiversity needs an integrated strategy. *Nature* **585**, 551–556 (2020).
5. J. M. Cavender-Bares, *et al.*, The hidden value of trees: Quantifying the ecosystem services of tree lineages and their major threats across the contiguous US. *PLOS Sustain Transform* **1**, e0000010 (2022).
6. J. A. Westfall, T. A. Schroeder, J. M. McCollum, P. L. Patterson, A spatial and temporal assessment of nonresponse in the national forest inventory of the U.S. *Environ Monit Assess* **194**, 530 (2022).
7. W. Turner, Sensing biodiversity. *Science* **346**, 301–302 (2014).
8. W. Turner, *et al.*, Free and open-access satellite data are key to biodiversity conservation. *Biological Conservation* **182**, 173–176 (2015).
9. W. Jetz, *et al.*, Monitoring plant functional diversity from space. *Nature Plants* **2**, 16024 (2016).
10. J.-B. Féret, G. P. Asner, Mapping tropical forest canopy diversity using high-fidelity imaging spectroscopy. *Ecological Applications* **24**, 1289–1296 (2014).
11. A. K. Schweiger, *et al.*, Plant spectral diversity integrates functional and phylogenetic components of biodiversity and predicts ecosystem function. *Nat Ecol Evol* **2**, 976–982 (2018).
12. R. Wang, *et al.*, Influence of species richness, evenness, and composition on optical diversity: A simulation study. *Remote Sensing of Environment* **211**, 218–228 (2018).
13. H. Gholizadeh, *et al.*, Detecting prairie biodiversity with airborne remote sensing. *Remote Sensing of Environment* **221**, 38–49 (2019).
14. H. Gholizadeh, J. A. Gamon, C. J. Helzer, J. Cavender-Bares, Multi-temporal assessment of grassland α- and β-diversity using hyperspectral imaging. *Ecological Applications* **30**, e02145 (2020).
15. F. C. Draper, *et al.*, Imaging spectroscopy predicts variable distance decay across contrasting Amazonian tree communities. *Journal of Ecology* **107**, 696–710 (2019).
16. A. K. Schweiger, E. Laliberté, Plant beta-diversity across biomes captured by imaging spectroscopy. *Nat Commun* **13**, 2767 (2022).
17. C. Rossi, *et al.*, Spatial resolution, spectral metrics and biomass are key aspects in estimating plant species richness from spectral diversity in species-rich grasslands. *Remote Sens Ecol Conserv* **8**, 297–314 (2022).

18. C. I. B. Wallis, *et al.*, Exploring the spectral variation hypothesis for α- and β-diversity: a comparison of open vegetation and forests. *Environ. Res. Lett.* **19**, 064005 (2024).
19. J. N. Pinto-Ledezma, A. K. Schweiger, J. A. Guzmán Q., J. Cavender-Bares, Plant diversity across dimensions: Coupling biodiversity measures from the ground and the sky. *Sci. Adv.* **11**, eadr0278 (2025).
20. A. K. Skidmore, *et al.*, Priority list of biodiversity metrics to observe from space. *Nat Ecol Evol* **5**, 896–906 (2021).
21. D. Rocchini, Distance decay in spectral space in analysing ecosystem β-diversity. *International Journal of Remote Sensing* **28**, 2635–2644 (2007).
22. D. Rocchini, K. S. He, J. Oldeland, D. Wesuls, M. Neteler, Spectral variation versus species β-diversity at different spatial scales: a test in African highland savannas. *J. Environ. Monit.* **12**, 825 (2010).
23. J. Dalmayne, T. Möckel, H. C. Prentice, B. C. Schmid, K. Hall, Assessment of fine-scale plant species beta diversity using WorldView-2 satellite spectral dissimilarity. *Ecological Informatics* **18**, 1–9 (2013).
24. J. J. Everest, *et al.*, Evaluating the utility of hyperspectral data to monitor local-scale β-diversity across space and time. *Remote Sensing of Environment* **316**, 114507 (2025).
25. A. Baselga, Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography* **19**, 134–143 (2010).
26. P. Cardoso, *et al.*, Partitioning taxon, phylogenetic and functional beta diversity into replacement and richness difference components. *Journal of Biogeography* **41**, 749–761 (2014).
27. J. N. Pinto-Ledezma, D. J. Larkin, J. Cavender-Bares, Patterns of Beta Diversity of Vascular Plants and Their Correspondence With Biome Boundaries Across North America. *Front. Ecol. Evol.* **6**, 194 (2018).
28. G. M. Moulatlet, *et al.*, Global patterns of phylogenetic beta-diversity components in angiosperms. *J Vegetation Science* **34**, e13203 (2023).
29. G. E. Hutchinson, *An introduction to population ecology* (Yale University Press, 1978).
30. J. Silvertown, Plant coexistence and the niche. *Trends in Ecology & Evolution* **19**, 605–611 (2004).
31. W. W. Hargrove, F. M. Hoffman, Using multivariate clustering to characterize ecoregion borders. *Comput. Sci. Eng.* **1**, 18–25 (1999).
32. D. K. Bolton, *et al.*, Continental-scale land surface phenology from harmonized Landsat 8 and Sentinel-2 imagery. *Remote Sensing of Environment* **240**, 111685 (2020).
33. B.-G. J. Brooks, D. C. Lee, L. Y. Pomara, W. W. Hargrove, Monitoring Broadscale Vegetational Diversity and Change across North American Landscapes Using Land Surface Phenology. *Forests* **11**, 606 (2020).
34. USDA Forest Service, Forest Inventory and Analysis Database. Deposited 2023.
35. W. A. Bechtold, P. L. Patterson, "The Enhanced Forest Inventory and Analysis Program - National Sampling Design and Estimation Procedures" (U.S. Department of Agriculture, Forest Service, Southern Research Station, 2015).
36. S. A. Smith, J. W. Brown, Constructing a broadly inclusive seed plant phylogeny. *American J of Botany* **105**, 302–314 (2018).
37. Y. Jin, H. Qian, V.PhyloMaker2: An updated and enlarged R package that can generate very large phylogenies for vascular plants. *Plant Diversity* **44**, 335–339 (2022).
38. R Core Team, R: A Language and Environment for Statistical Computing. (2023). Deposited 2023.
39. P. Cardoso, F. Rigal, J. C. Carvalho, BAT – Biodiversity Assessment Tools, an R package for the measurement and estimation of alpha and beta taxon, phylogenetic and functional diversity. *Methods Ecol Evol* **6**, 232–236 (2015).
40. S. E. Fick, R. J. Hijmans, WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Intl Journal of Climatology* **37**, 4302–4315 (2017).
41. J. Oksanen, *et al.*, vegan: Community Ecology Package. (2025). Deposited 2025.
42. D. Krutz, *et al.*, The Instrument Design of the DLR Earth Sensing Imaging Spectrometer (DESIS). *Sensors* **19**, 1622 (2019).

43. D. R. Thompson, *et al.*, On-orbit calibration and performance of the EMIT imaging spectrometer. *Remote Sensing of Environment* **303**, 113986 (2024).
44. B. Rivard, J. Feng, A. Gallie, A. Sanchez-Azofeifa, Continuous wavelets for the improved use of spectral libraries and hyperspectral data. *Remote Sensing of Environment* **112**, 2850–2862 (2008).
45. F. A. Kruse, *et al.*, The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. *Remote Sensing of Environment* **44**, 145–163 (1993).
46. J. A. Knott, *et al.*, Effects of outliers on remote sensing-assisted forest biomass estimation: A case study from the United States national forest inventory. *Methods Ecol Evol* **14**, 1587–1602 (2023).
47. R. E. McRoberts, Q. Chen, B. F. Walters, D. J. Kaisershot, The effects of global positioning system receiver accuracy on airborne laser scanning-assisted estimates of aboveground biomass. *Remote Sensing of Environment* **207**, 42–49 (2018).
48. M. Kuhn, K. Johnson, *Applied predictive modeling*, 1st Ed. (Springer, 2013).
49. H. Meyer, C. Reudenbach, T. Hengl, M. Katurji, T. Nauss, Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. *Environmental Modelling & Software* **101**, 1–9 (2018).
50. H. Gu, A. Singh, P. A. Townsend, Detection of gradients of forest composition in an urban area using imaging spectroscopy. *Remote Sensing of Environment* **167**, 168–180 (2015).
51. N. Lunardon, G. Menardi, N. Torelli, ROSE: a Package for Binary Imbalanced Learning. *The R Journal* **6**, 79 (2014).
52. J. A. Guzmán Q., *et al.*, Mapping oak wilt disease from space using land surface phenology. *Remote Sensing of Environment* **298**, 113794 (2023).

Supporting Information

Supplementary methods

Methods S1. Selection of inventoried plots from the Forest Inventory and Analysis (FIA) program across the Eastern U.S.

The FIA program currently conducts inventories using an annualized inventory system, collecting data on a subset of plots each year, generally 1/7 to 1/5 of plots in the Eastern U.S., leading to a 5- or 7-year inventory cycle. We selected FIA plots from Eastern U.S. states by FIA evaluation ID (EVALID) with inventory cycles ending in 2018-2022, the most recent inventories available as of the data query on 27 April 2023 (Table S1). We limited our query to plots with at least one forested condition. From these plots, we selected all live trees > 12.7 cm diameter at breast height (DBH) that are sampled on four circular 7.31 m radius subplots. Tree diameter and height were used to estimate wood volume assuming a cylinder shape, and volume was converted to abundance per unit area by multiplying by FIA trees-per-acre adjustment factor (TPA_UNADJ), which is based on the ratio of the subplot sampling area to one acre of land. We used the National Ecological Observatory Network (NEON) ecoregion map (1) to further filter plots from forested ecoregions.

Methods S2. Integration of plant functional traits.

We collected data for eight plant functional traits: tree maximum height (H), tree slenderness (S), wood density (WD), shade tolerance index (STI), drought tolerance index (DTI), waterlogging tolerance index (WTI), leaf mass per area (LMA), and nitrogen content (N). H (m) was estimated as the 95-quantile value of tree height per species among all the FIA plots that were used. S was estimated as the 50-quantile value of the ratio between tree height and diameter at the breast height for all the trees per species among all the FIA plots that were used. Wood density (g m^{-3}) was obtained directly from the FIA database per each species. STI, DTI, and WTI were obtained from (2). LMA (g m^{-2}) and N (mg g^{-1}) were obtained from the Botanical Information and Ecology Network (BIEN) database through R (3), accessed on April 6, 2024. Data were missing for some species for traits gathered from (2) and BIEN, and therefore, we imputed them using a phylogenetic approach following (4). We used a pruned phylogenetic tree from (5) to generate a phylogenetic distance matrix of ten orthogonal eigenvectors using the 'PVRdecomp' function of the PVR package of R (6). All plant traits with missing values and the ten orthogonal eigenvectors were integrated into the 'missForest' function (7) using 500 iterations and 100 trees to impute missing values. The resulting imputed values were used to fill in missing traits (e.g., Fig. S25).

We applied a principal components analysis (PCA) to the complete set of plant traits per species to summarize the variation of traits among species into three principal components. The first, second, and third principal components explained 24.28, 17.88, and 16.32% of the variation in plant traits (Fig. S26), respectively. We then extracted PCA scores of the first three principal components per species to construct a functional trait dendrogram using euclidean distances among species. Finally, the functional trait dendrogram was transformed into a phylogenetic tree class in R using the "as.phylogeny" command in APE (8) to generate functional dissimilarity values for the calculation of functional beta diversity.

Methods S3. Estimation of spectral dissimilarity and its regression with dimensions of beta diversity.

We compared pairwise spectral dissimilarity values among communities on selected clear-sky spaceborne observations of FIA plots. These pairwise comparisons were calculated using smoothed reflectance spectra, not on wavelet transformations of spectra. To smooth the reflectance spectra, we applied a Savitzky-Golay smoothing filter on the extracted pixels from the FIA plots. This filter was applied using a first order and two neighboring bands of length from the target bands (i.e., $n = 5$). The application of the Savitzky-Golay filter was done using the 'sgolayfilt' function of the signal package in R (9). We then removed bands on the resulting smoothed spectra to match spectral regions used in our modelling efforts (ie., DESIS: 449 – 965 nm; EMIT: 447 – 1265, 1488 – 1710, and 2015 – 2424 nm).

Methods S4. Climatic and topographic datasets.

We used well-established climatic and topographic datasets at 1 km of spatial resolution to rotate MDS β-diversity axes and predict forest attributes. Specifically, we employed historic climatic datasets for mean annual temperature (MAT), temperature seasonality (TS), temperature annual range (TAR), mean annual precipitation (MAP), and precipitation seasonality (PS) derived from Worldclim 2 (10). Likewise, we employed topographic datasets associated with the elevation (ELE), slope (SL), and compound topographic index (CTI) (e.g., wetness index) derived from HYDRO1k (11). The extraction of pixels from these datasets were conducted using publicly available coordinates of FIA plots given their coarse resolution. To integrate these climatic and topographic datasets with scenes of forest community composition to map forest attributes, we extracted values using the centroid location of pixels from spaceborne observations.

Supplementary tables

Table S1. Inventory end years, states, and FIA evaluation IDs (EVALIDs) included in this study.

Inventory end year	States	EVALIDs
2018	Kentucky	211801
2019	Florida, Louisiana, Massachusetts, New York, Ohio, Oklahoma, Tennessee	121901, 221901, 251901, 361901, 391901, 401901, 471901
2020	Kansas, Maryland, Michigan, Minnesota, New Jersey, Vermont, West Virginia	202001, 242001, 262001, 272001, 342001, 502001, 542001
2021	Arkansas, Connecticut, Delaware, Georgia, Illinois, Indiana, Maine, Mississippi, Missouri, New Hampshire, North Carolina, Pennsylvania, Rhode Island, South Carolina, Texas, Virginia, Wisconsin	52101, 92101, 102101, 132101, 172101, 182101, 232101, 282101, 292101, 332101, 372101, 422101, 442101, 452101, 482121, 512101, 552101
2022	Alabama	12201

Table S2. Performance summary of repeated Partial Least Square Regression models using spaceborne observations of DESIS to predict Multi-Dimensional Scaling (MDS) axes based on matrices of dimensions of beta diversity. Performance metrics are described by the coefficient of determination (R^2), bias, the Root Mean Square Error (RMSE), and the percentage of RMSE (%RMSE) based on the data range. Values represent the mean and standard deviation of 100 models.

Dimension	Dataset	Axes	Performance metric			
			R^2	BIAS (x10^{-5})	RMSE	%RMSE
Taxonomic	Training	Axis 1	0.38 ± 0.00	-6.9 ± 51.71	0.17 ± 0.00	21.84 ± 0.01
		Axis 2	0.14 ± 0.00	-3.3 ± 33.15	0.15 ± 0.00	25.29 ± 0.01
		Axis 3	0.08 ± 0.00	14.6 ± 20.27	0.12 ± 0.00	18.25 ± 0.00
	Testing	Axis 1	0.36 ± 0.00	402.7 ± 53.60	0.16 ± 0.00	21.87 ± 0.02
		Axis 2	0.11 ± 0.00	-320.6 ± 36.32	0.16 ± 0.00	25.92 ± 0.02
		Axis 3	0.07 ± 0.00	377.6 ± 23.62	0.12 ± 0.00	18.90 ± 0.01
Phylogenetic	Training	Axis 1	0.17 ± 0.00	2.2 ± 32.86	0.14 ± 0.00	20.70 ± 0.01
		Axis 2	0.20 ± 0.00	15 ± 43.89	0.19 ± 0.00	20.33 ± 0.01
		Axis 3	0.33 ± 0.00	-7.3 ± 72.08	0.26 ± 0.00	27.15 ± 0.01
	Testing	Axis 1	0.15 ± 0.00	-156.7 ± 36.13	0.14 ± 0.00	21.29 ± 0.02
		Axis 2	0.17 ± 0.00	631.3 ± 45.19	0.18 ± 0.00	20.33 ± 0.01
		Axis 3	0.31 ± 0.00	1265.3 ± 72.83	0.26 ± 0.00	27.36 ± 0.02
Functional	Training	Axis 1	0.30 ± 0.00	0.4 ± 48.43	0.19 ± 0.00	23.28 ± 0.01
		Axis 2	0.16 ± 0.00	19.3 ± 30.56	0.16 ± 0.00	20.03 ± 0.00
		Axis 3	0.30 ± 0.00	13.8 ± 60.67	0.20 ± 0.00	22.08 ± 0.01
	Testing	Axis 1	0.27 ± 0.00	112.3 ± 51.48	0.19 ± 0.00	23.31 ± 0.02
		Axis 2	0.13 ± 0.00	378 ± 31.04	0.15 ± 0.00	20.25 ± 0.01
		Axis 3	0.29 ± 0.00	731.1 ± 63.07	0.20 ± 0.00	22.51 ± 0.01

Table S3. Performance summary of repeated Partial Least Square Regression models using spaceborne observations of EMIT to predict Multi-Dimensional Scaling (MDS) axes based on matrices of dimensions of beta diversity. Performance metrics are described by the coefficient of determination (R^2), bias, the Root Mean Square Error (RMSE), and the percentage of RMSE (%RMSE) based on the data range. Values represent the mean and standard deviation of 100 models.

Dimension	Evaluation	Axes	Performance metric			
			R^2	BIAS (x10^{-5})	RMSE	%RMSE
Taxonomic	Training	Axis 1	0.64 ± 0.00	-3.90 ± 58.41	0.13 ± 0.00	17.07 ± 0.01
		Axis 2	0.14 ± 0.00	-1.70 ± 28.04	0.14 ± 0.00	23.84 ± 0.01
		Axis 3	0.33 ± 0.00	32.50 ± 54.22	0.17 ± 0.00	18.88 ± 0.01
	Testing	Axis 1	0.61 ± 0.00	212.30 ± 59.32	0.13 ± 0.00	17.40 ± 0.02
		Axis 2	0.13 ± 0.00	-290.50 ± 31.28	0.15 ± 0.00	24.30 ± 0.01
		Axis 3	0.31 ± 0.00	96.80 ± 55.58	0.18 ± 0.00	19.60 ± 0.02
Phylogenetic	Training	Axis 1	0.27 ± 0.00	2.90 ± 34.71	0.12 ± 0.00	19.56 ± 0.01
		Axis 2	0.10 ± 0.00	20.40 ± 32.47	0.19 ± 0.00	21.24 ± 0.00
		Axis 3	0.27 ± 0.00	7.90 ± 74.02	0.29 ± 0.00	29.46 ± 0.01
	Testing	Axis 1	0.24 ± 0.00	-578.90 ± 36.46	0.13 ± 0.00	20.55 ± 0.02
		Axis 2	0.08 ± 0.00	349.00 ± 32.36	0.19 ± 0.00	20.67 ± 0.01
		Axis 3	0.24 ± 0.00	483.00 ± 77.06	0.29 ± 0.00	29.63 ± 0.02
Functional	Training	Axis 1	0.50 ± 0.00	-3.60 ± 59.64	0.17 ± 0.00	19.49 ± 0.01
		Axis 2	0.07 ± 0.00	13.40 ± 24.05	0.15 ± 0.00	20.55 ± 0.01
		Axis 3	0.31 ± 0.00	10.40 ± 65.01	0.23 ± 0.00	24.07 ± 0.01
	Testing	Axis 1	0.46 ± 0.00	261.90 ± 60.66	0.17 ± 0.00	20.15 ± 0.02
		Axis 2	0.05 ± 0.00	229.30 ± 25.63	0.15 ± 0.00	20.39 ± 0.02
		Axis 3	0.30 ± 0.00	32.10 ± 64.04	0.23 ± 0.00	24.47 ± 0.02

Table S4. Detailed list of spectral regions, its sensitivity, and potential ecological meaning with influence for predicting MDS β-diversity axes of multiple dimensions of diversity using spaceborne observations of DESIS and EMIT.

Wavelengths (nm)	Biophysical/biochemical sensitivity	Potential meaning associated with plant communities
553	Reflectance peak due to low absorption by chlorophyll	Proxy for variation in chlorophyll content and canopy health among communities
710	Chlorophyll absorption shoulder	Photosynthetic capacity and leaf nitrogen content among communities
720	Start of the near red-edge inflection	Sensitive to chlorophyll and canopy stress; early sign of senescence
756	Near red-edge inflection	Strong in vegetation indices; indicates species differences
790	Canopy structure	Sensitive to variations in leaf area index, canopy density, and leaf scattering properties among communities
973	Water absorption	Community-level water content
1116	Leaf structural traits and dry matter	Associated with dry matter, LMA, and some lignin/cellulose content
1161	Continuation of dry matter region	Tracks cell structure, leaf dry matter; functionally differentiates among communities
1521	Strong water absorption	Community-level water content
2226	Dry matter, protein, cellulose	Strong absorption by lignin, cellulose, protein; reflects C:N ratio, functional type, successional status

Supplementary figures

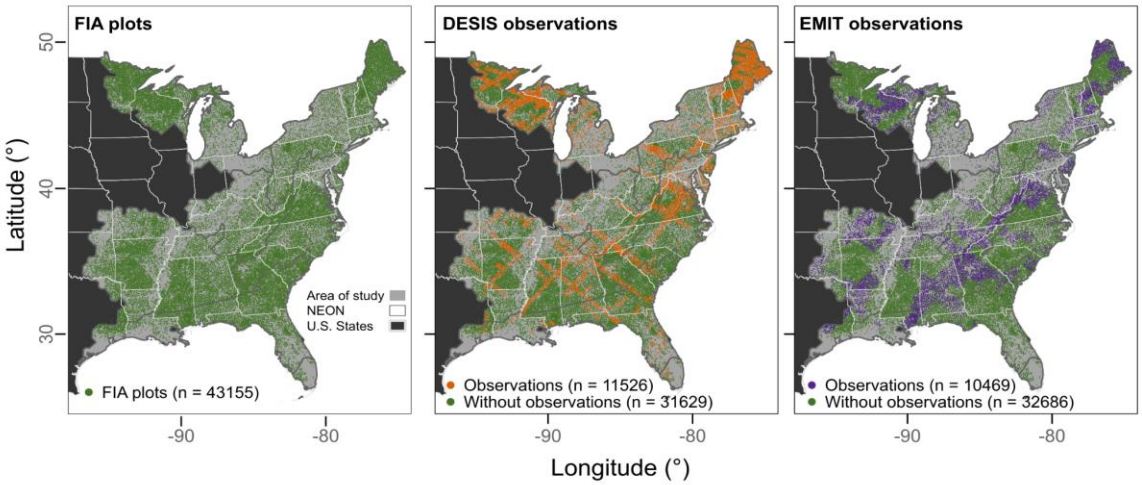


Fig S1. Spatial distribution of blurred plot locations from the Forest Inventory Analysis program and the availability of clear sky observations from DESIS and EMIT spaceborne observations.

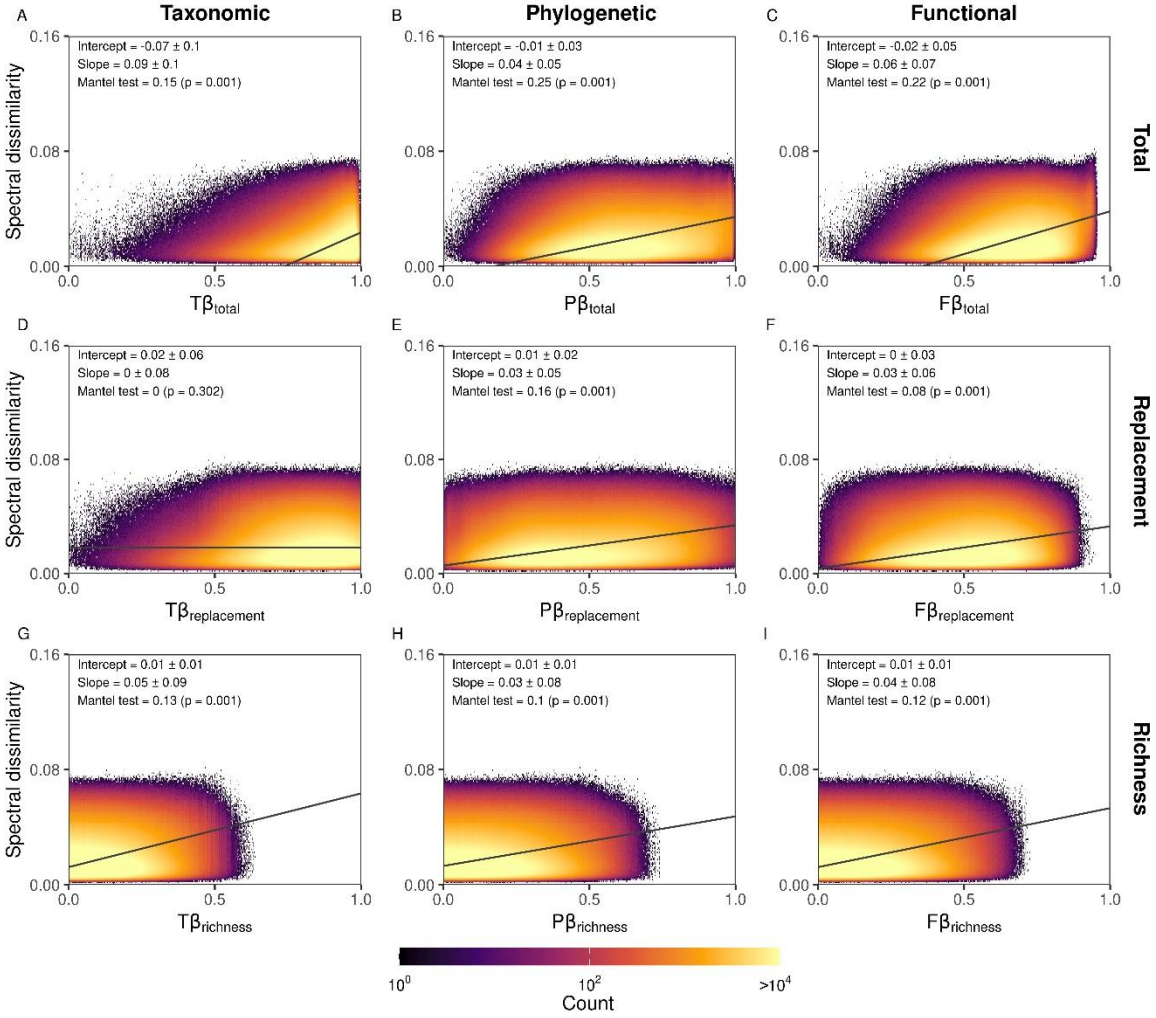


Fig S2. Relationships between beta diversity (β) and spectral dissimilarity from DESIS observations among communities. Horizontal panels describe the dimensions of β diversity, including taxonomic (Tβ), phylogenetic (Pβ), and functional (Fβ) diversity. Vertical panels describe the partition of beta total diversity (β_{total}) into replacement ($\beta_{replacement}$) and richness ($\beta_{richness}$). The black solid lines represent the average linear regression line when regressing dissimilarities from a target community with all communities.

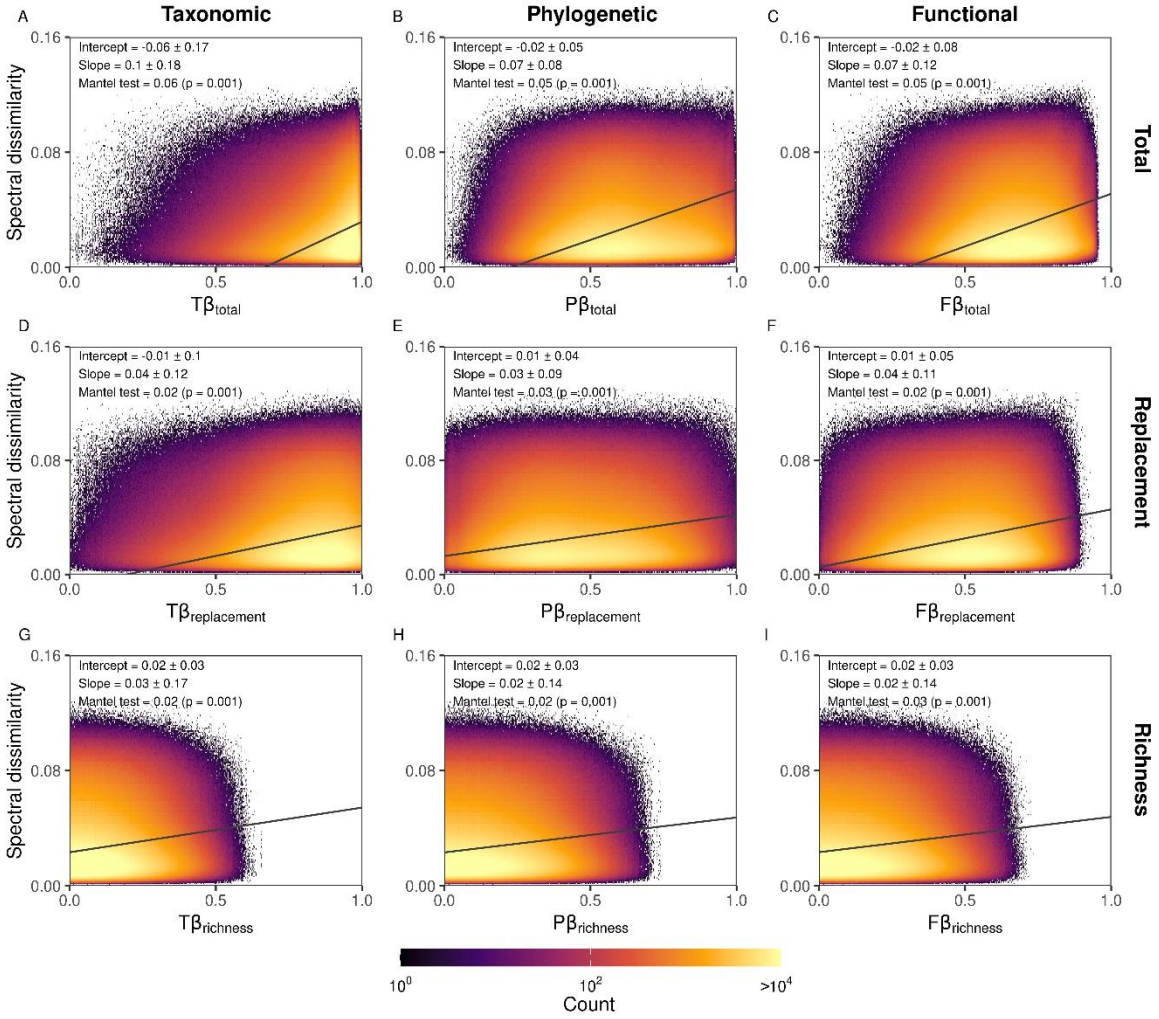


Fig S3. Relationship between beta diversity (β) and spectral dissimilarity from EMIT observations among communities. Horizontal panels describe the dimensions of β diversity, including taxonomic (Tβ), phylogenetic (Pβ), and functional (Fβ) diversity. Vertical panels describe the partition of beta total diversity (β_{total}) into replacement ($\beta_{replacement}$) and richness ($\beta_{richness}$). The black solid lines represent the average linear regression line when regressing dissimilarities from a target community with all communities.

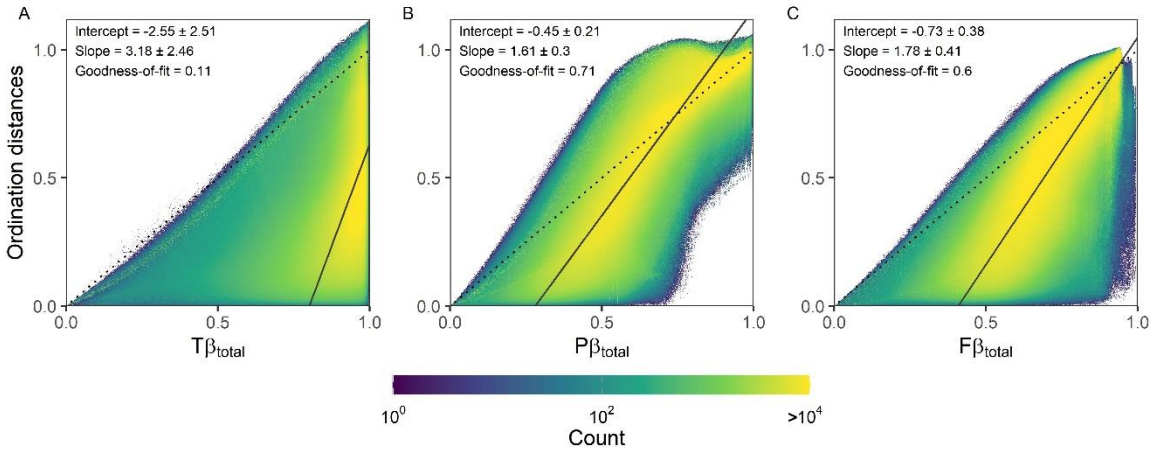


Fig. S4. Scatterplot between ordination distances derived from the Multi-Dimensional Scaling analyses and the dimensions of beta total diversity based on taxonomic (A), phylogenetic (B), and functional (C) information. The black solid lines represent the average linear regression line when regressing distances between a target community and all other communities. Dotted lines represent the 1:1 line. The goodness-of-fit is calculated from the coefficient of determination of correlations between ordination axes values and beta distances.

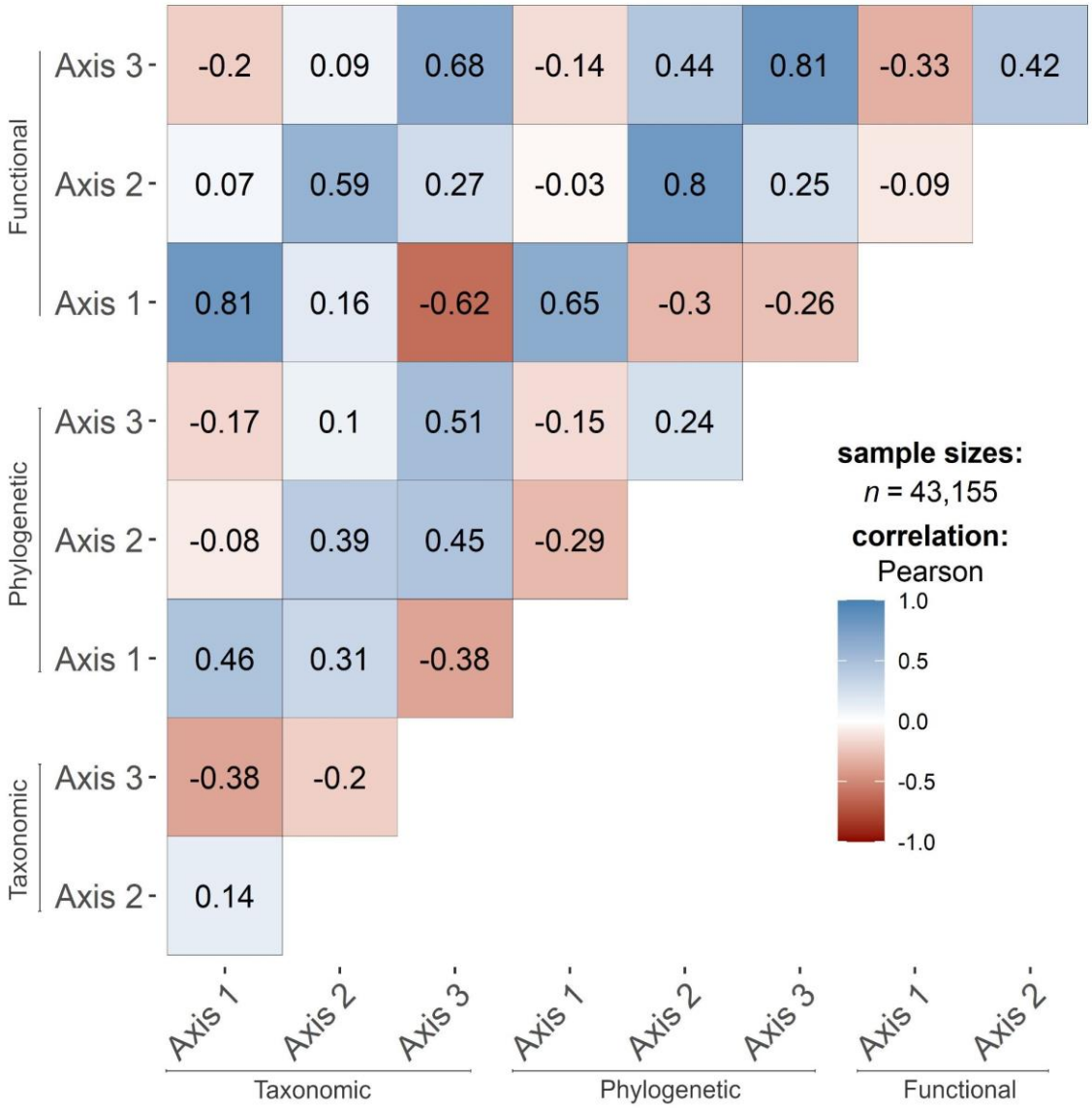


Fig S5. Correlation coefficients between MDS axes of ordination for three dimensions beta diversity, including taxonomic, phylogenetic, and functional beta diversity.

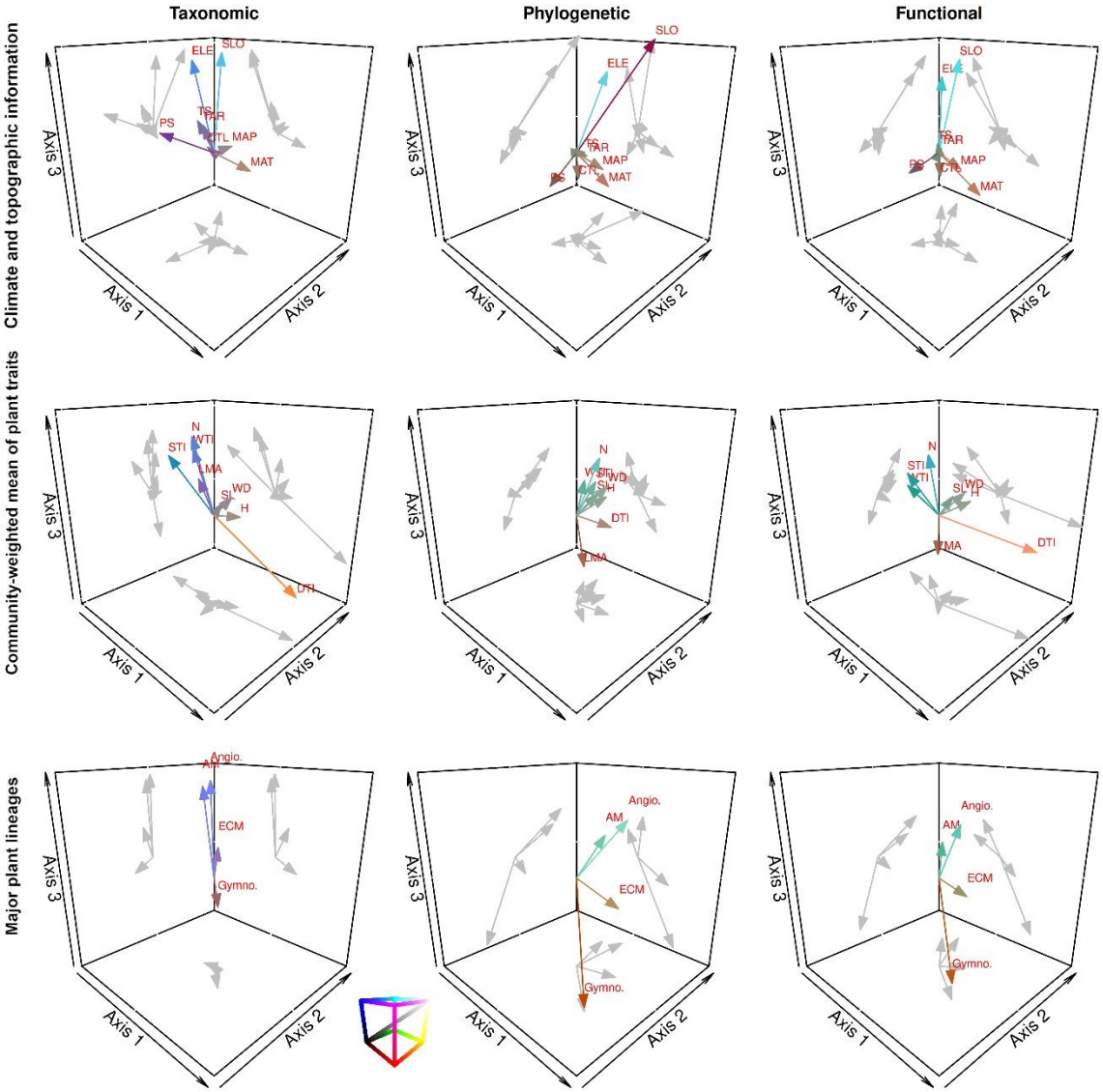


Fig S6. Three-dimensional projection of weighted average scores of climatic and topography characteristics, community-weighted mean of plant traits, and relative abundance of major plant lineages of communities within Multi-Dimensional Scaling ordinations derived from taxonomic, phylogenetic, and functional beta diversity. Acronyms represent, mean annual temperature (MAT), temperature seasonality (TS), temperature annual range (TAR), annual precipitation (MAP), precipitation seasonality (PS), elevation (ELE), slope (SLO), and compound topographic index (CTI), tree maximum height (H), tree slenderness (SL), wood density (WD), shade tolerance index (STI), drought tolerance index (DTI), waterlogging tolerance index (WTI), leaf mass per area (LMA), nitrogen content (N), gymnosperms (Gymno.), angiosperms (Angio.), ectomycorrhizal symbiosis (ECM), arbuscular mycorrhizal symbiosis (AM).

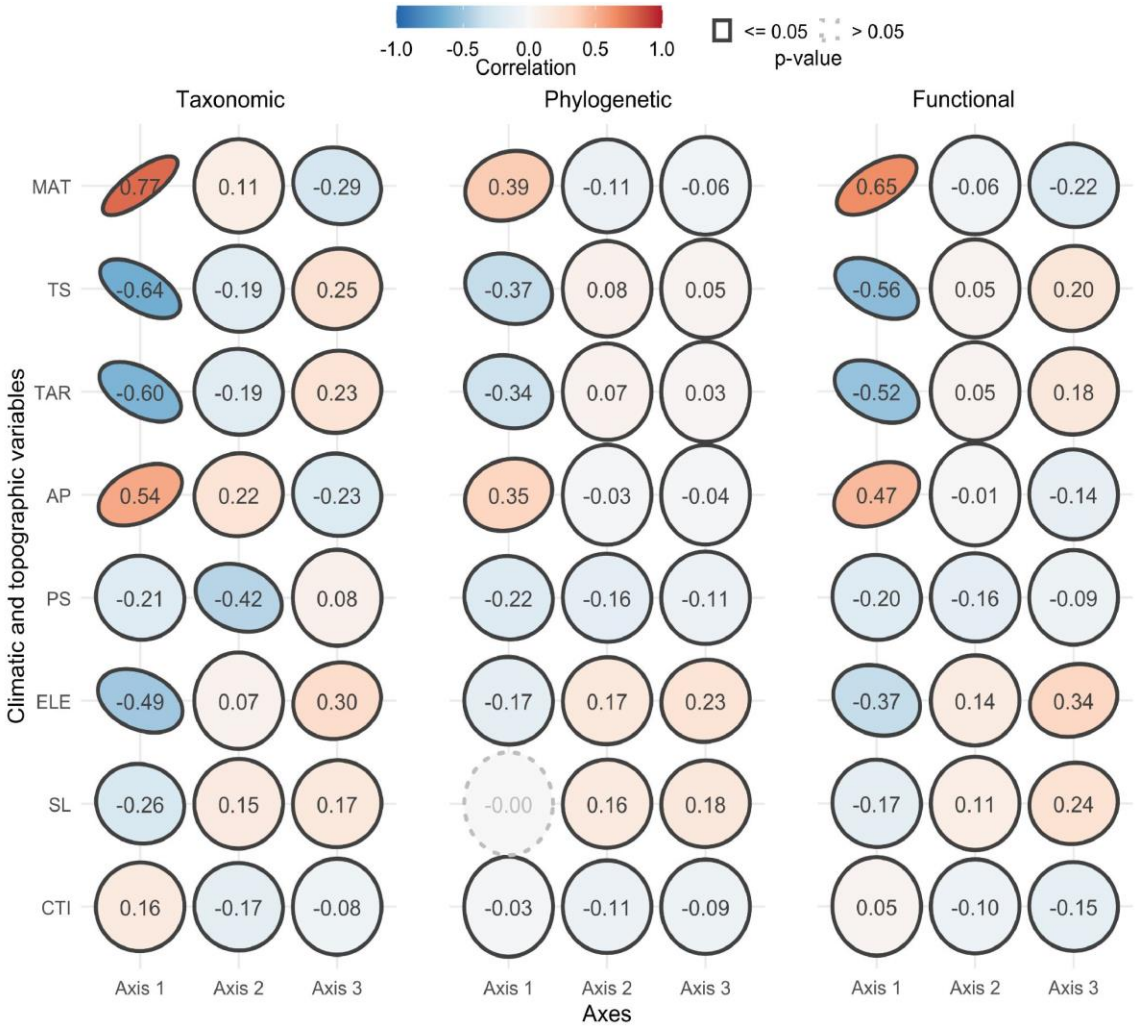


Fig S7. Correlation between axes of Multi-Dimensional Scaling ordinations derived from taxonomic, phylogenetic, and functional beta diversity and climatic and topographic variables of forest communities. Acronyms represent: mean annual temperature (MAT), temperature seasonality (TS), temperature annual range (TAR), annual precipitation (AP), precipitation seasonality (PS), elevation (ELE), slope (SL), and compound topographic index (CTI).

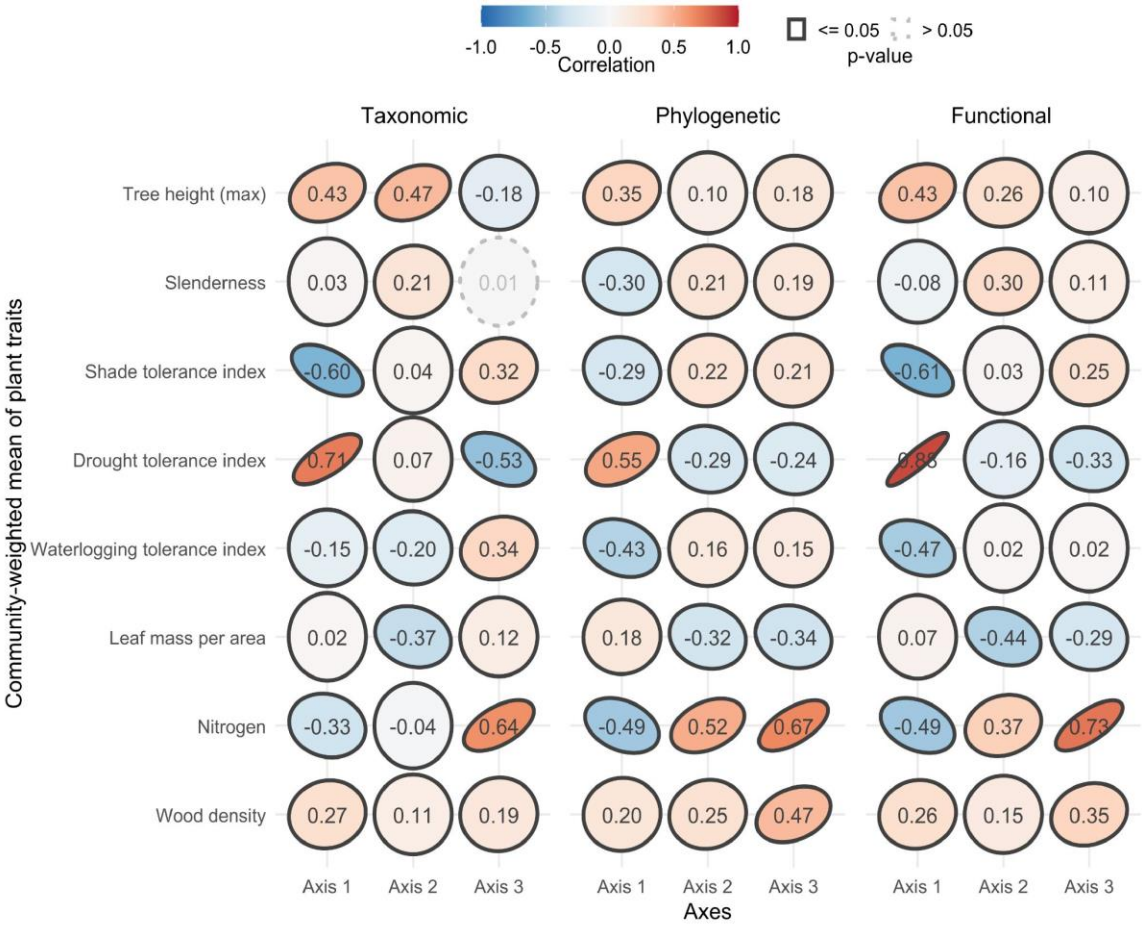


Fig S8. Correlation between axes of Multi-Dimensional Scaling ordinations derived from taxonomic, phylogenetic, and functional beta diversity and community-weighted mean of plant traits of forest communities.

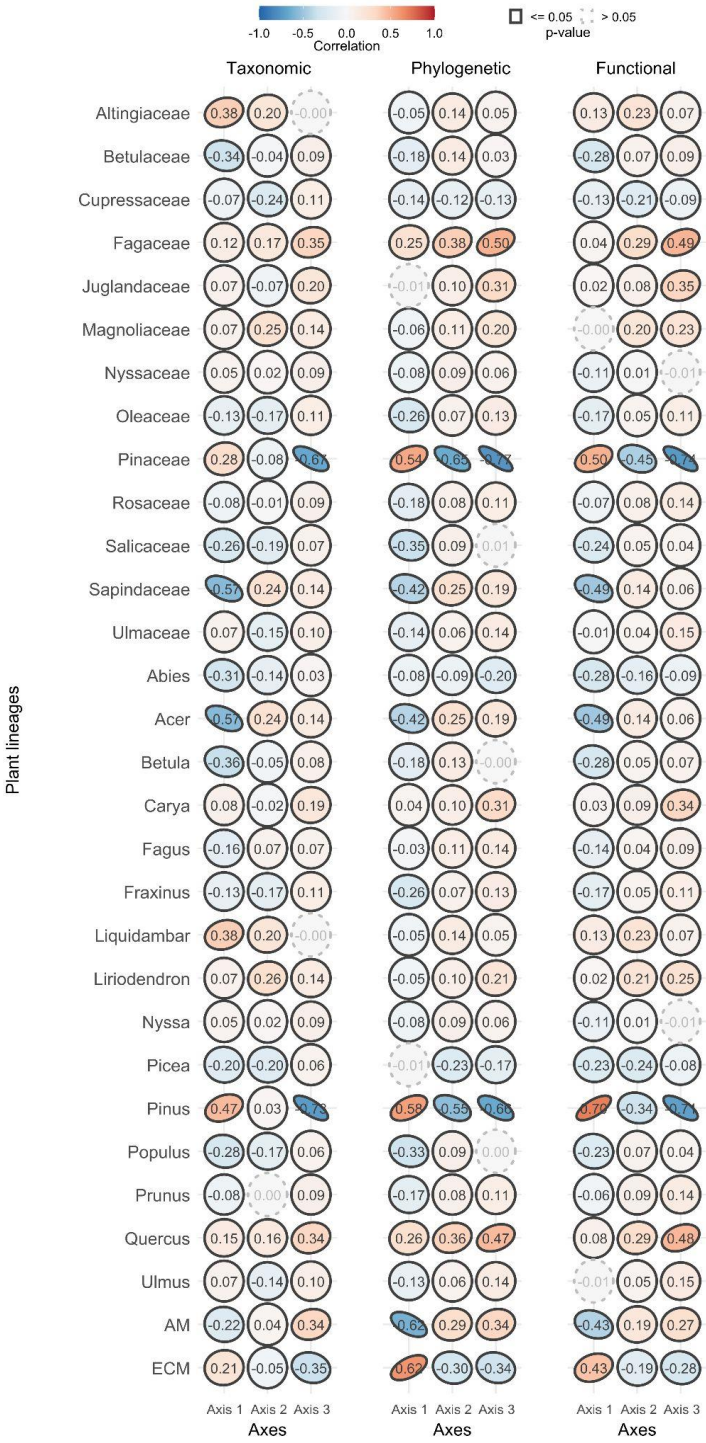


Fig S9. Correlation between axes of Multi-Dimensional Scaling ordinations derived from taxonomic, phylogenetic, and functional beta diversity and the relative abundance of plant lineages within forest communities.

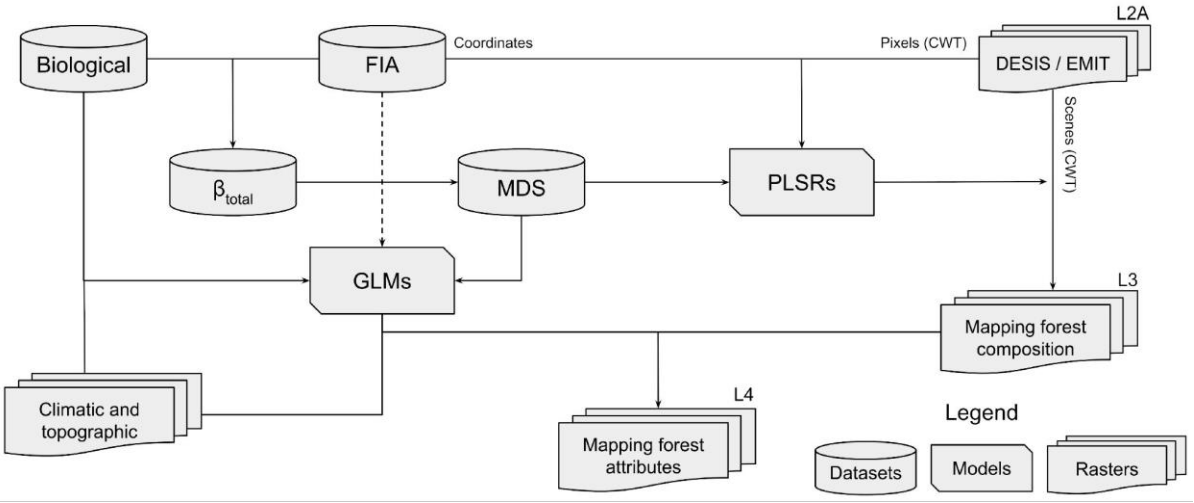
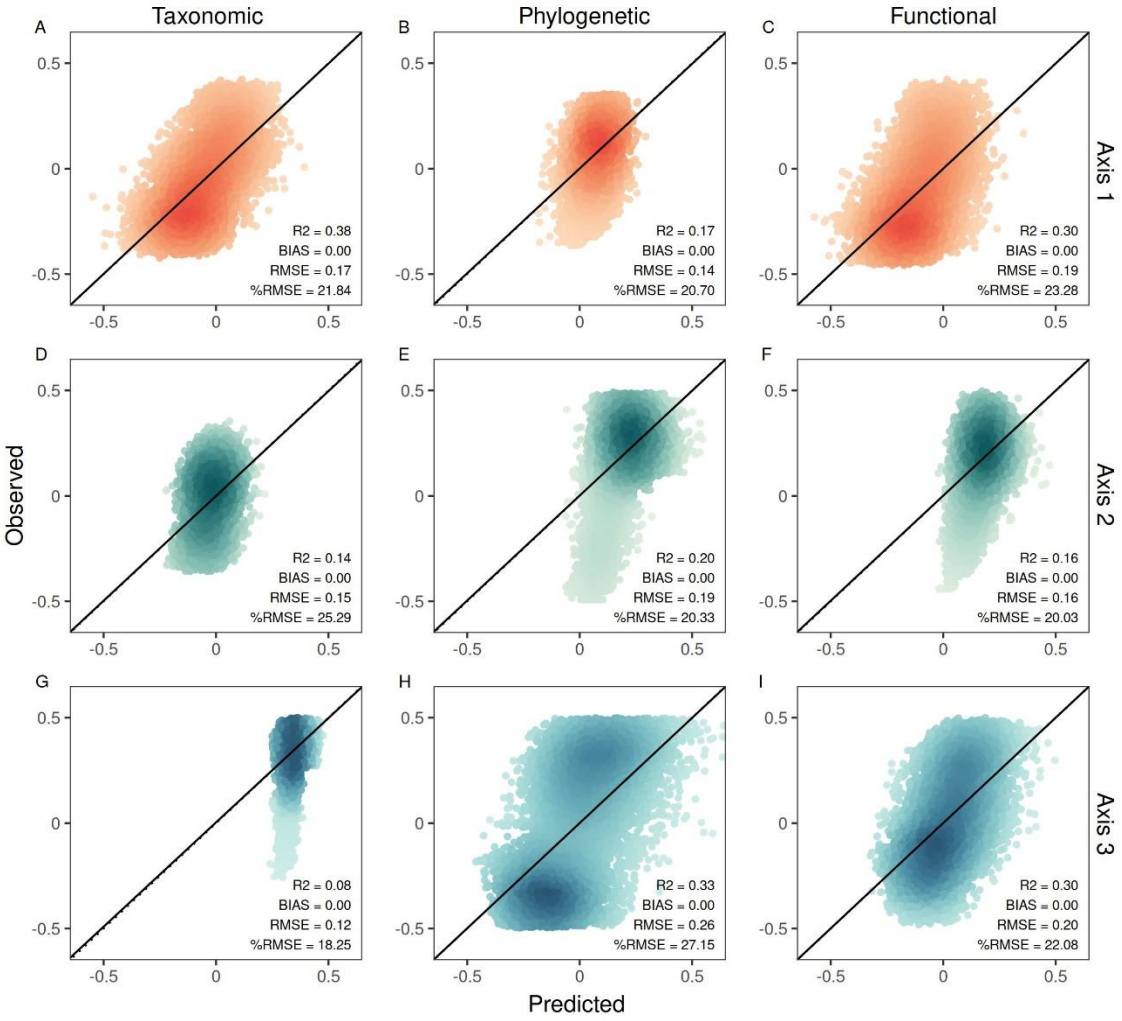


Fig. S10. Schematic representation of the workflow used to map multiple dimensions of forest diversity.



773
774 **Fig S11**. Performance of training datasets of PLSR models that predict ordination axes of multiple
775 dimensions of tree diversity using spaceborne observations from DESIS. Values represent the mean and
776 standard deviation of 100 models.

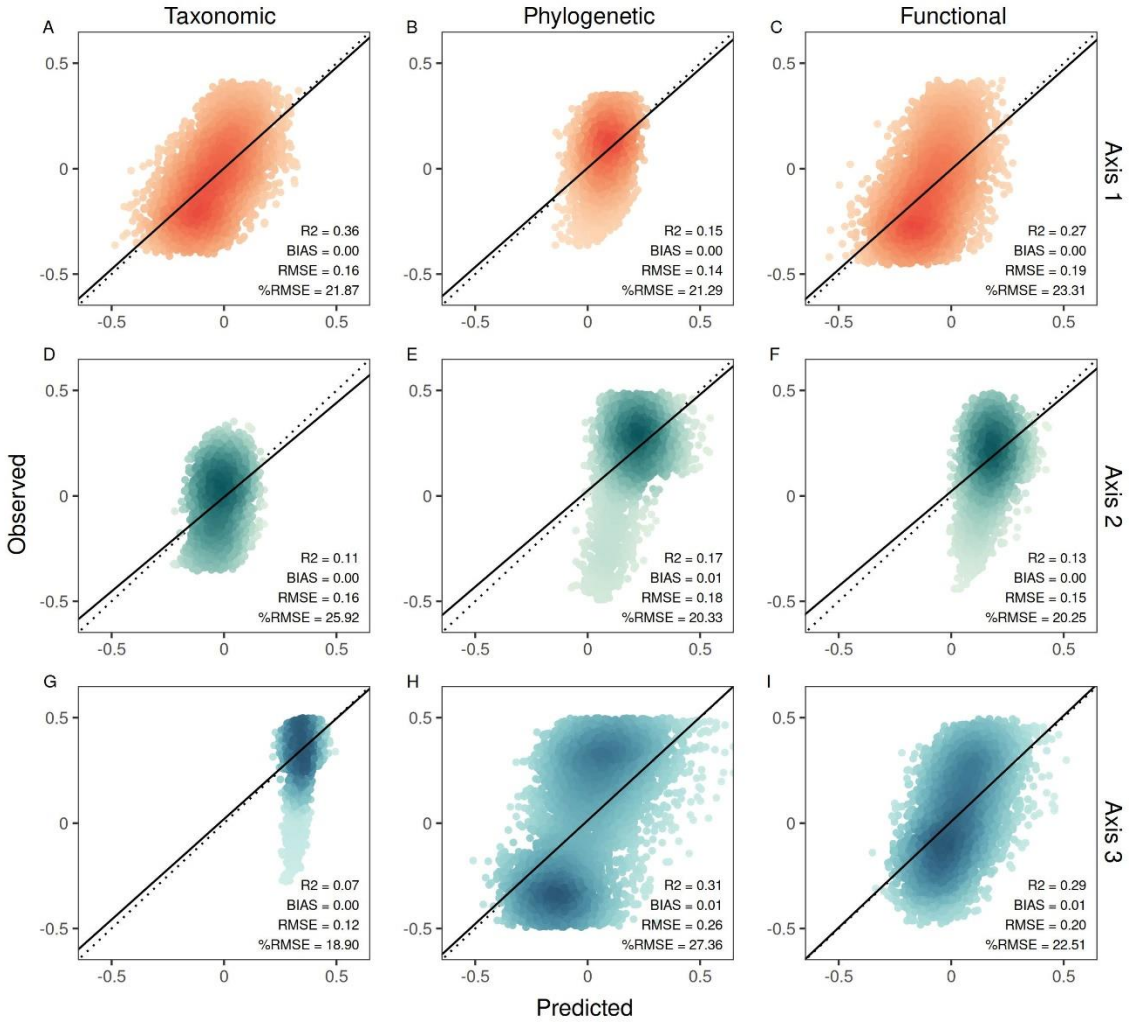


Fig 12. Performance of testing datasets of PLSR models that predict ordination axes of multiple dimensions of tree diversity using spaceborne observations from DESIS. Values represent the mean and standard deviation of 100 models.

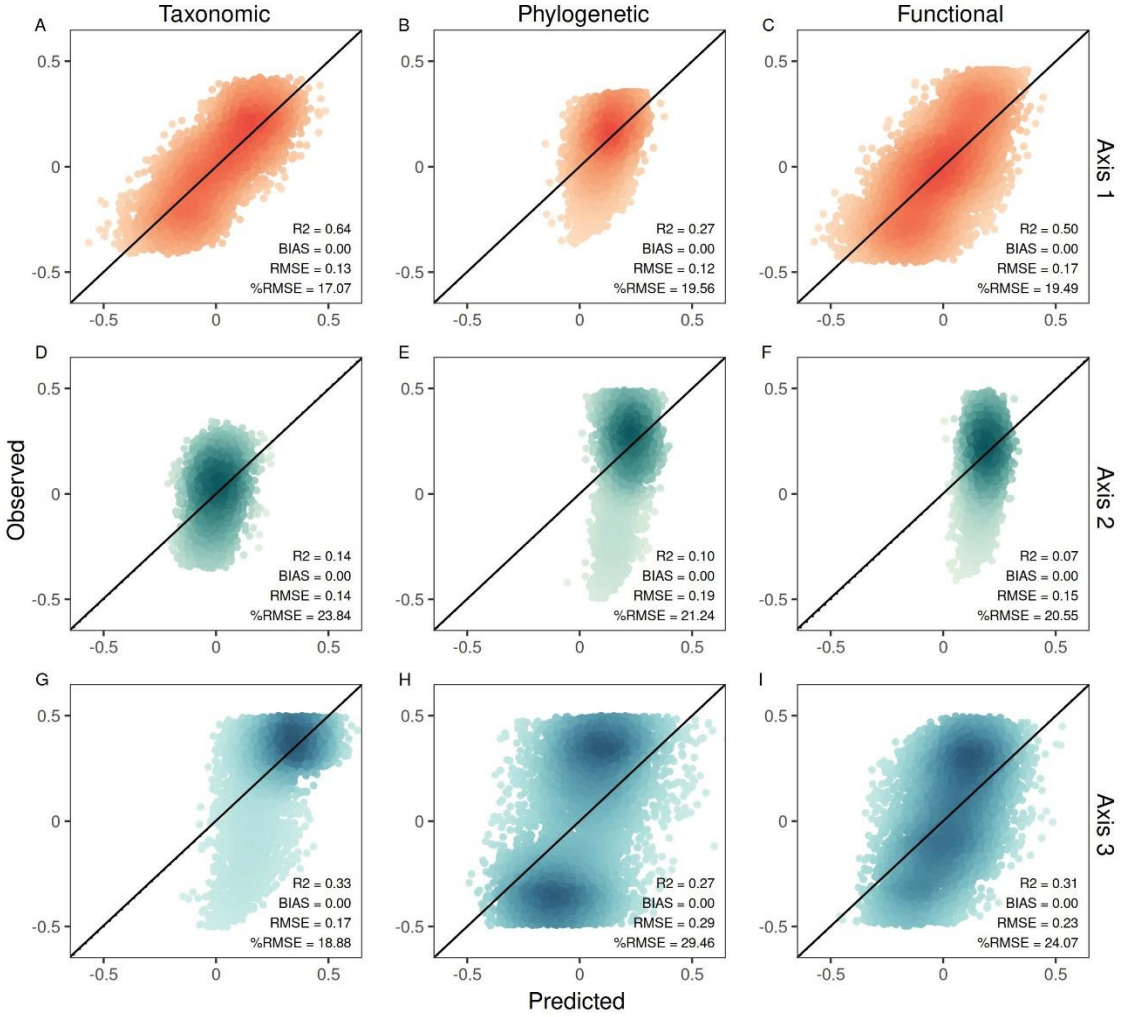


Fig S13. Performance of training datasets of PLSR models that predict ordination axes of multiple dimensions of tree diversity using spaceborne observations from EMIT. Values represent the mean and standard deviation of 100 models.

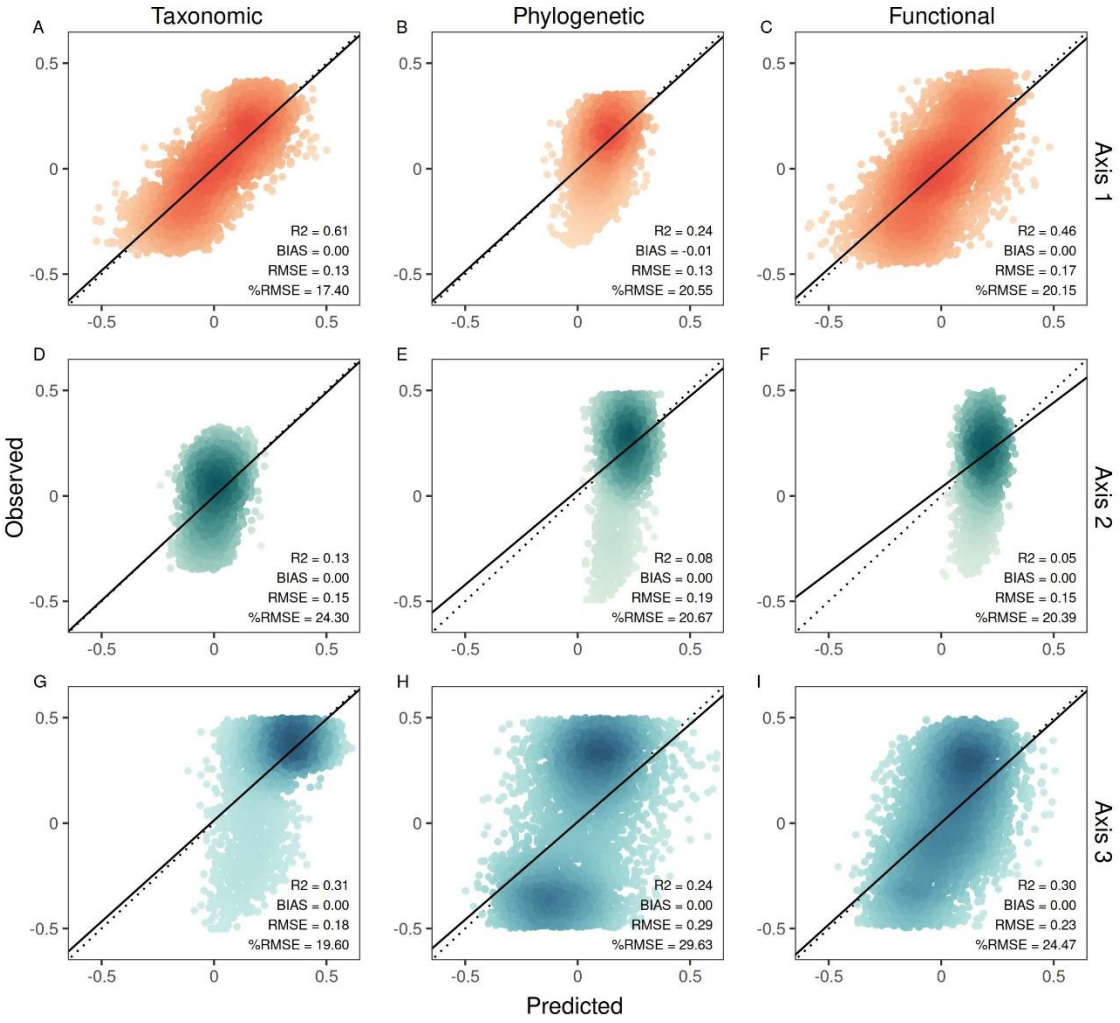


Fig S14. Performance of testing datasets of PLSR models that predict ordination axes of multiple dimensions of tree diversity using spaceborne observations from EMIT. Values represent the mean and standard deviation of 100 models.

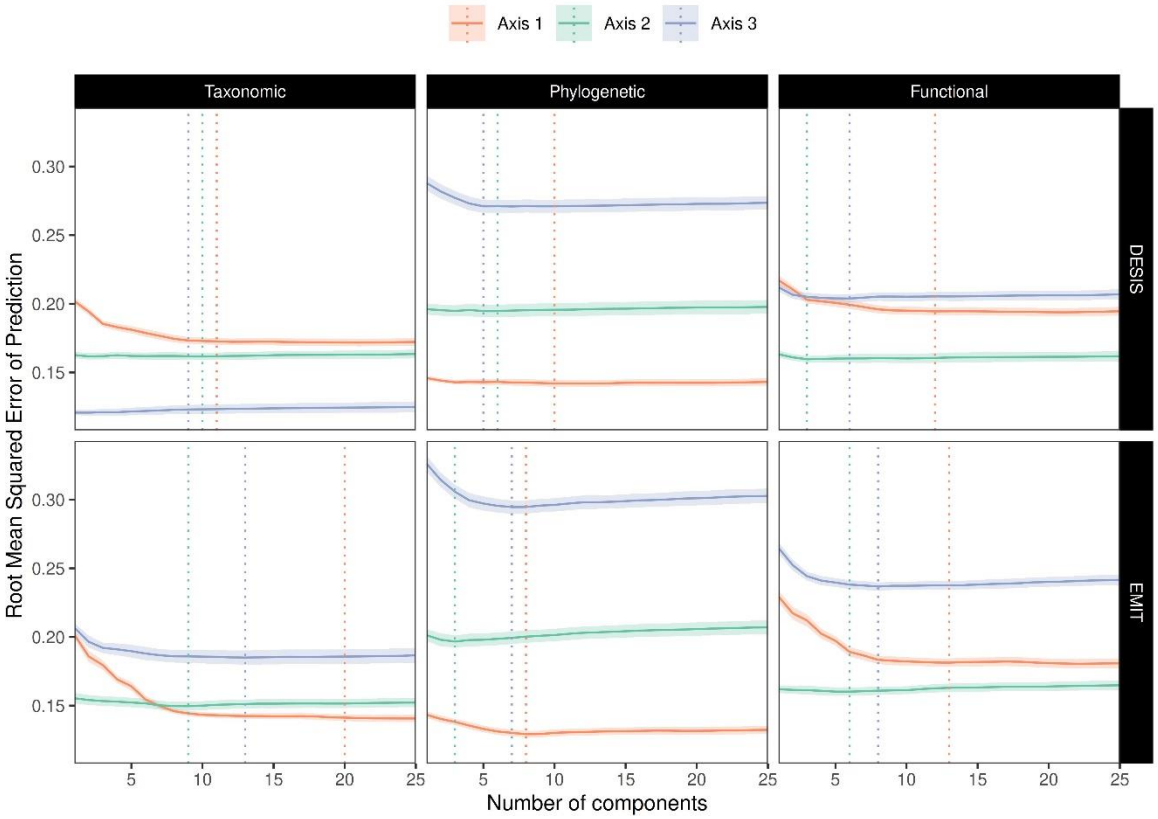


Fig. S15. Root Mean Squared Error of Prediction (RMSEP) from cross-validation derived from PLSR models that predict ordination axes of multiple dimensions of tree diversity using spaceborne observations from DESIS and EMIT. Each solid line represents the average of the 100 iterations, while the shade around each line is its standard deviation. Horizontal dotted lines represent the optimal number of components selected for each axis.

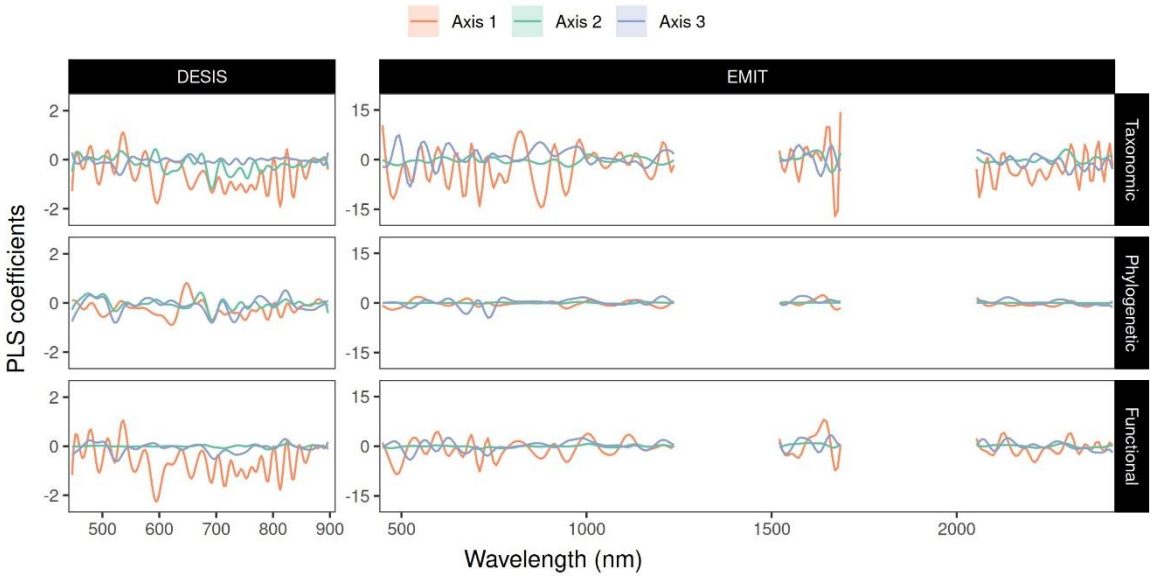


Fig S27. PLSR coefficients of the models that predict ordination axes of multiple dimensions of tree diversity using wavelet-transformed spaceborne reflectance from DESIS and EMIT. Each line represents an average of 100 iterations, while the shade around each line represents the standard deviation.

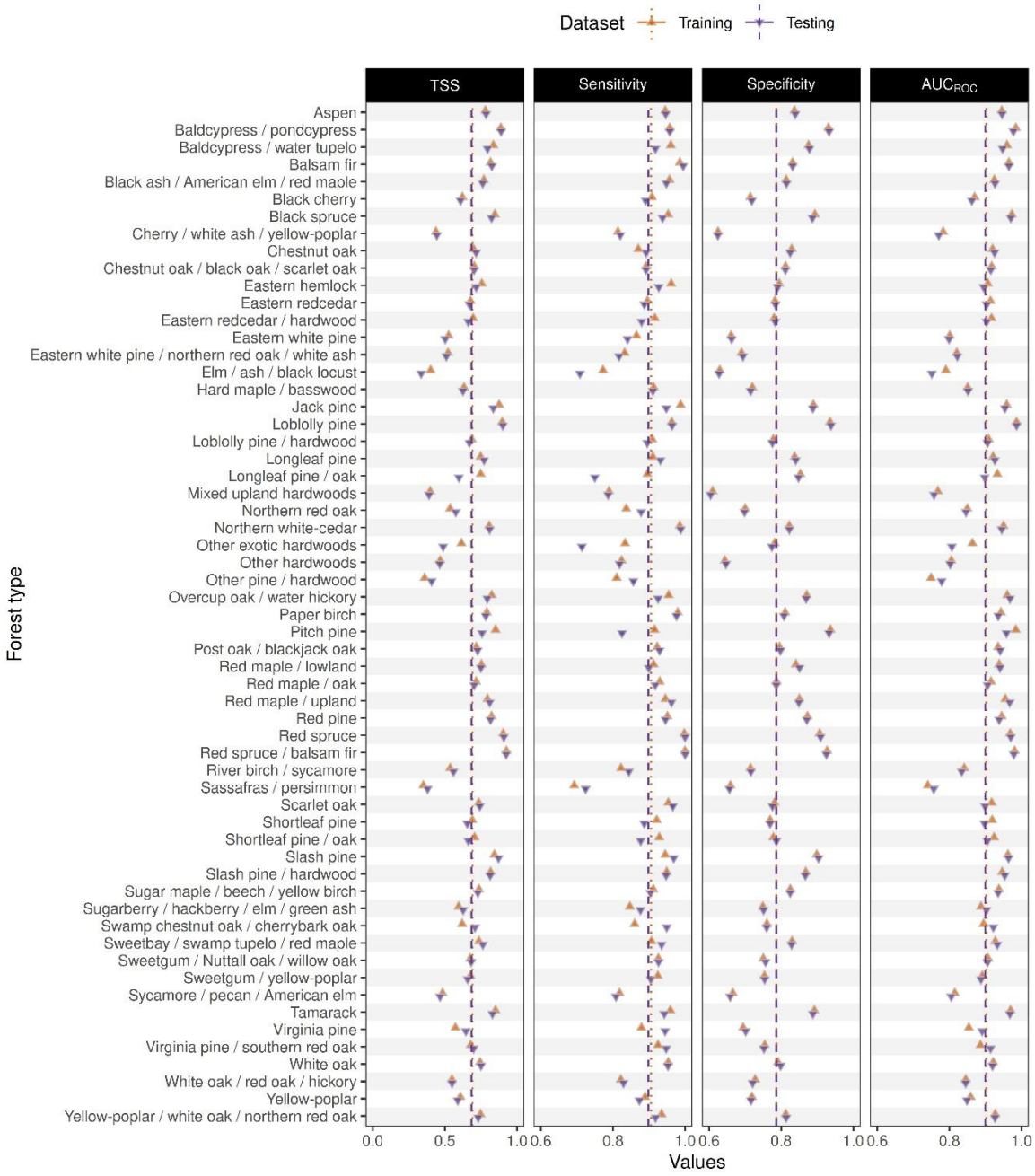


Fig S16. Performance of binary generalized linear models to predict labels of forest types from the Forest Inventory and Analysis program using ordination axes of beta diversity based on taxonomic information, climatic, and topographic data. Each point represents the mean estimate of 100 models.

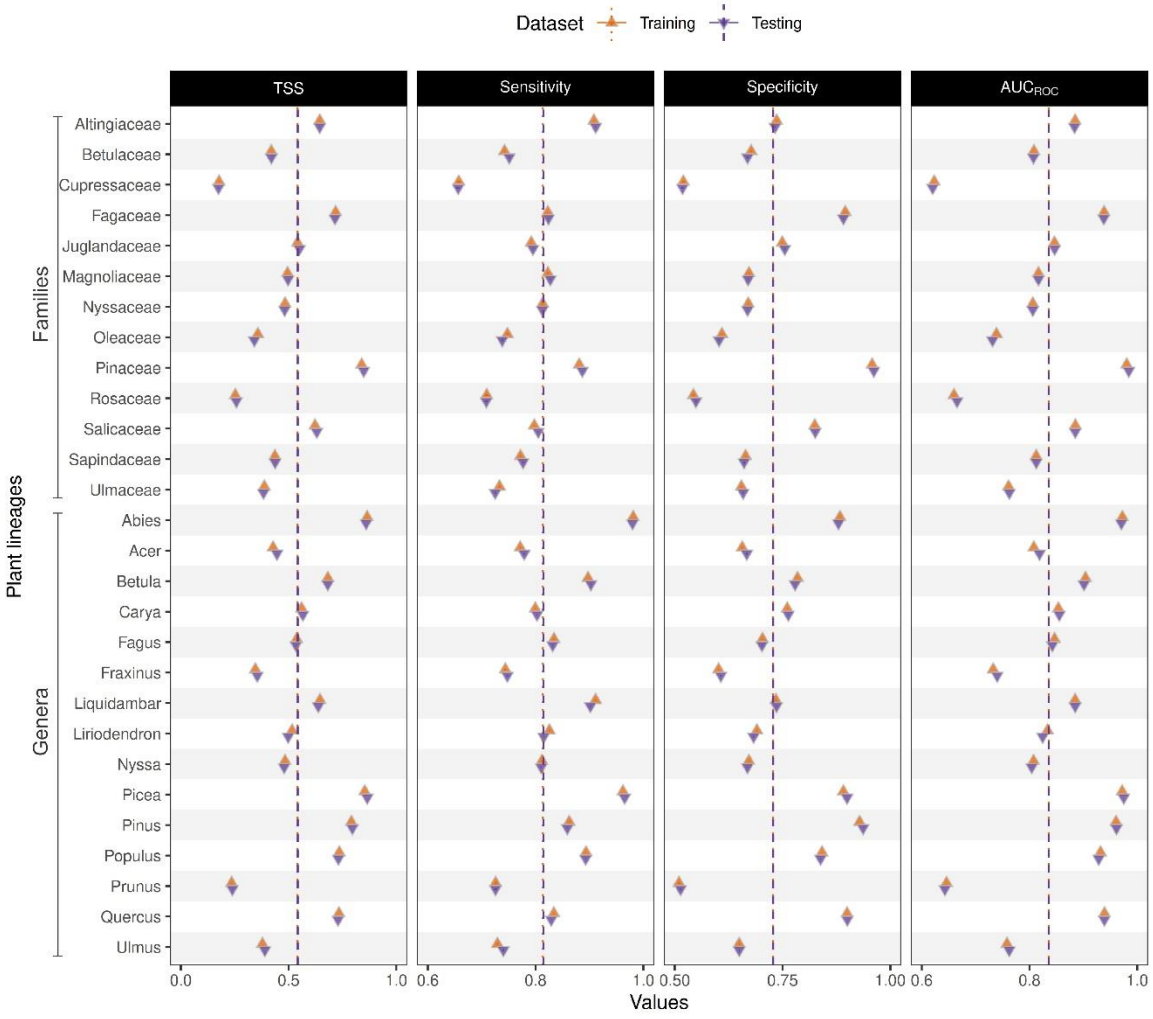


Fig S17. Performance of binary generalized linear models to predict the presence / absence of plant lineages within inventories from the Forest Inventory and Analysis program using ordination axes of beta diversity based on phylogenetic information, climatic, and topographic data. Each point represents the mean estimate of 100 models.

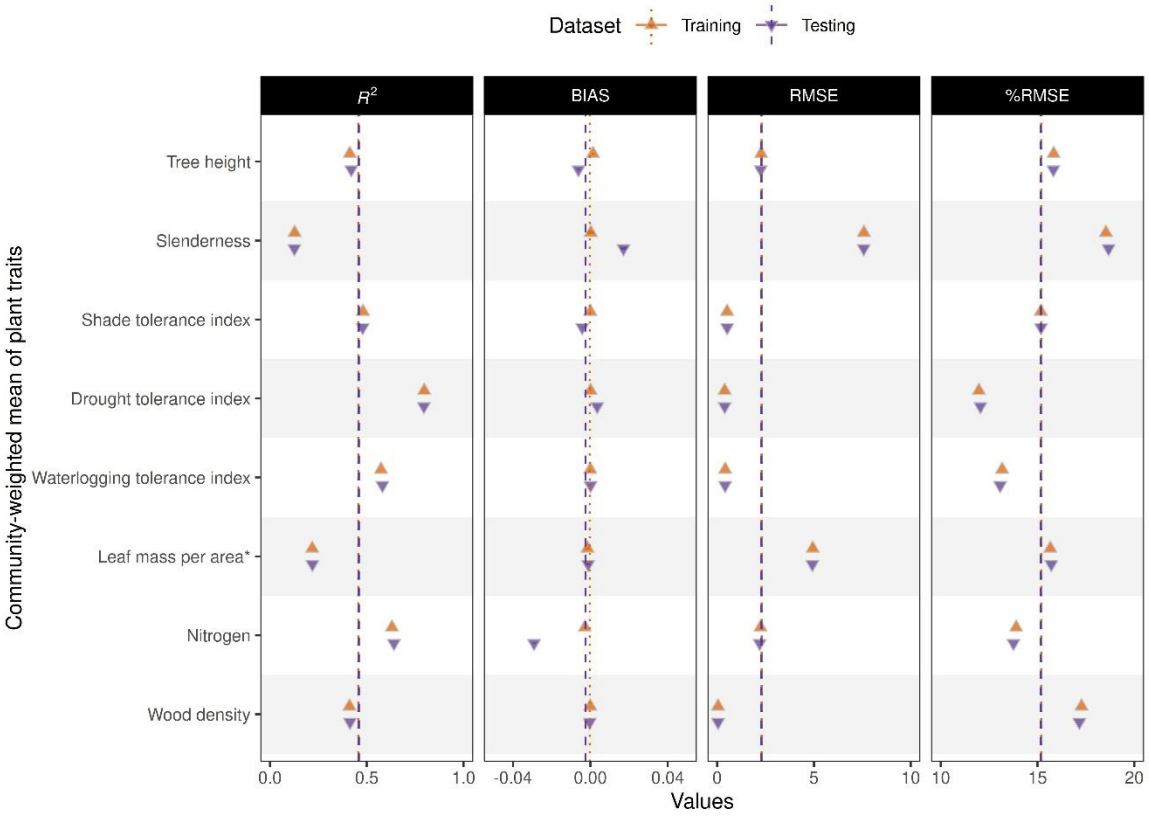


Fig S18. Performance of generalized linear models to predict community-weighted mean of plant traits within inventories from the Forest Inventory and Analysis program using ordination axes of beta diversity based on functional information, climatic, and topographic data. Each point represents the mean estimate of 100 models.

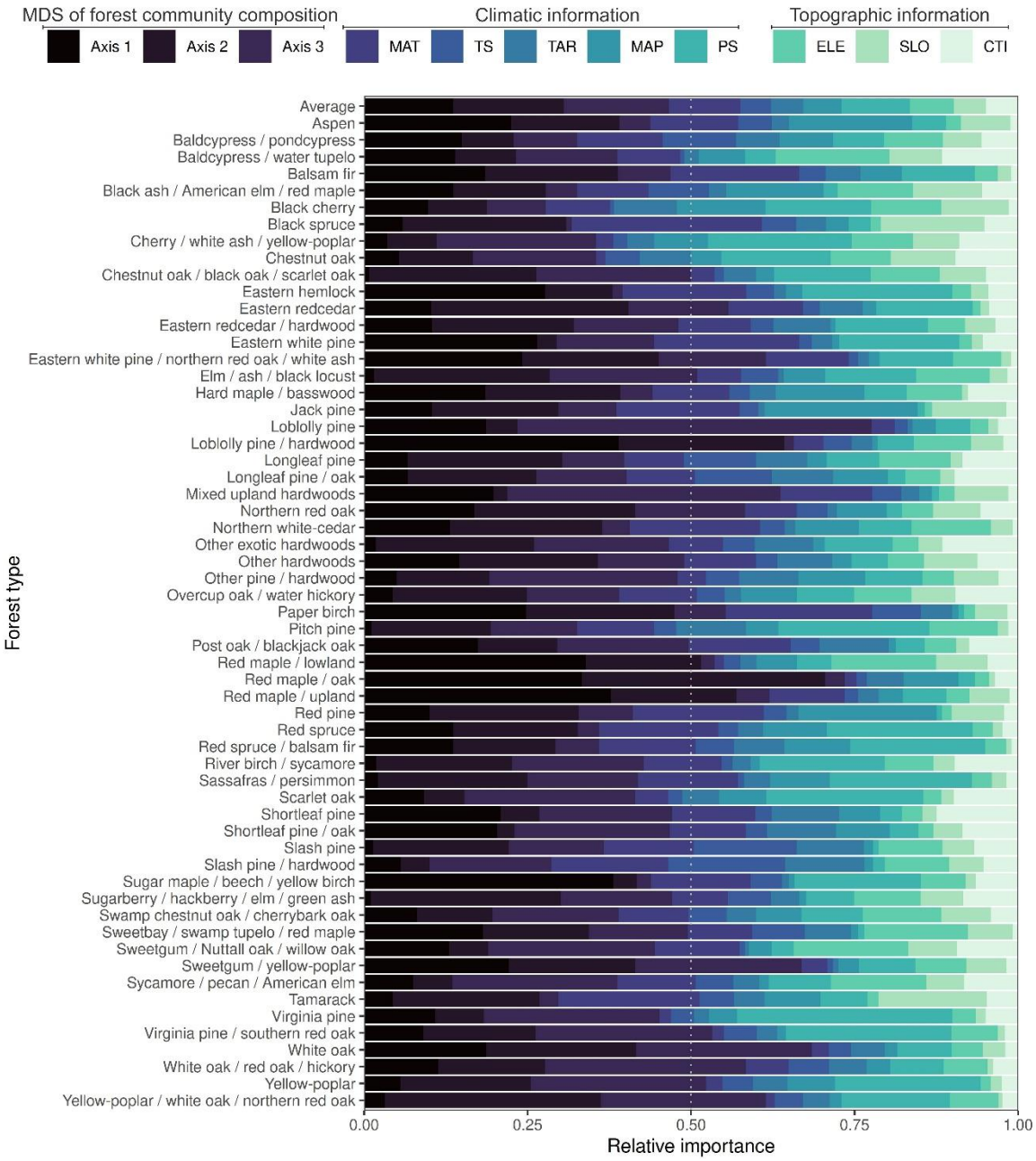


Fig S19. Relative importance of variables used in binary generalized linear models to predict presence/absence of forest types within the Forest Inventory and Analysis program through axes of forest composition, climatic, and topographic information. Axes (i.e., Axis 1, Axis 2, and Axis 3) describe the potential forest composition from a MDS ordination of beta diversity based on taxonomic information. Climatic variables are described by the mean annual temperature (MAT), the temperature seasonality (TS), temperature annual range (TAR), annual precipitation (AP), and precipitation seasonality (PS). Topographic information is described by elevation (ELE), slope (SL), and compound topographic index (CTI).

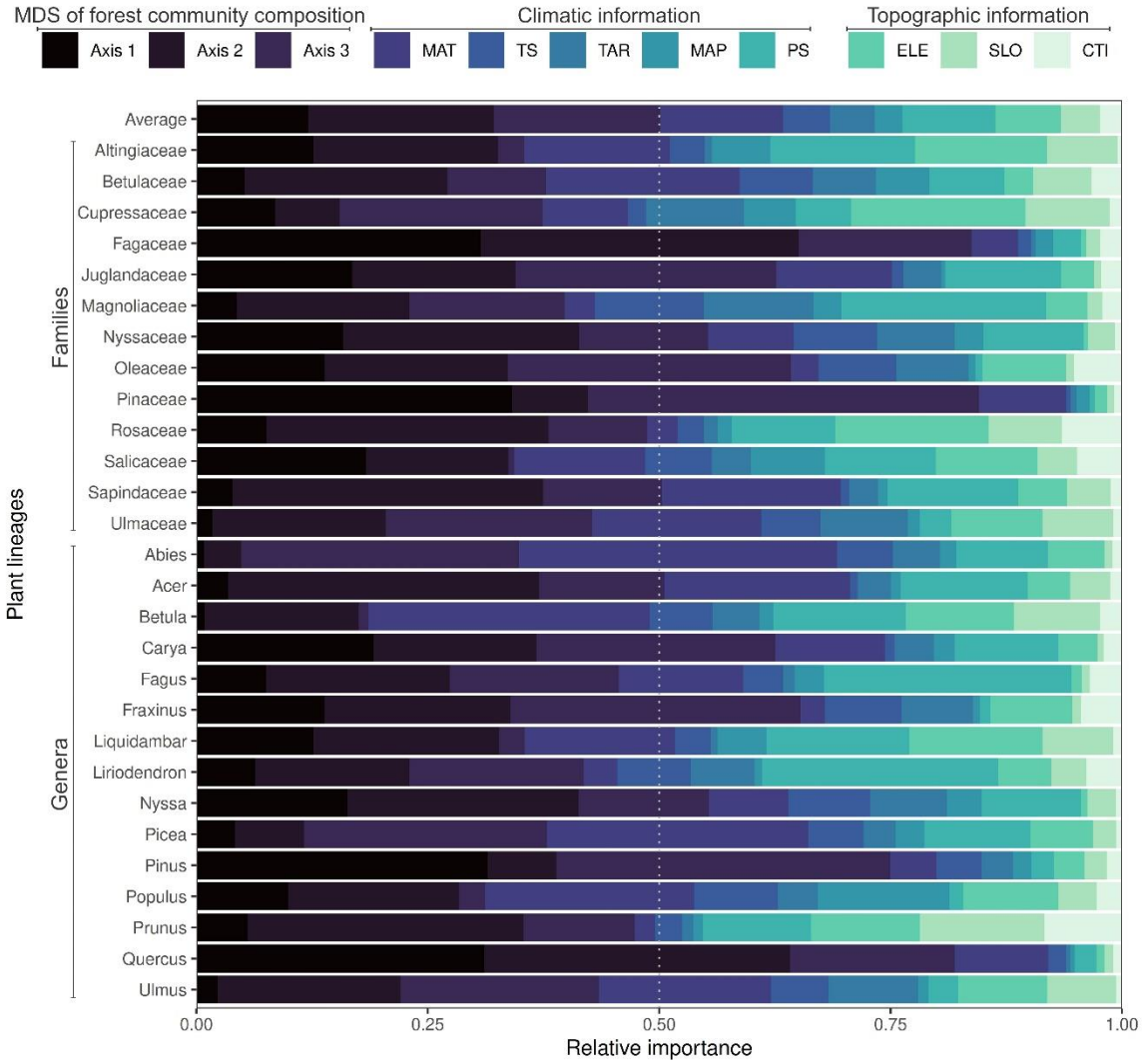


Fig S20. Relative importance of variables used in binary generalized linear models to predict presence/absence of plant lineages within inventories of the Forest Inventory and Analysis program through axes of forest composition, climatic, and topographic information. Axes (i.e., Axis 1, Axis 2, and Axis 3) describe the potential forest composition from a MDS ordination of beta diversity based on phylogenetic information. Climatic variables are described by the mean annual temperature (MAT), the temperature seasonality (TS), temperature annual range (TAR), annual precipitation (AP), and precipitation seasonality (PS). Topographic information is described by elevation (ELE), slope (SL), and compound topographic index (CTI).

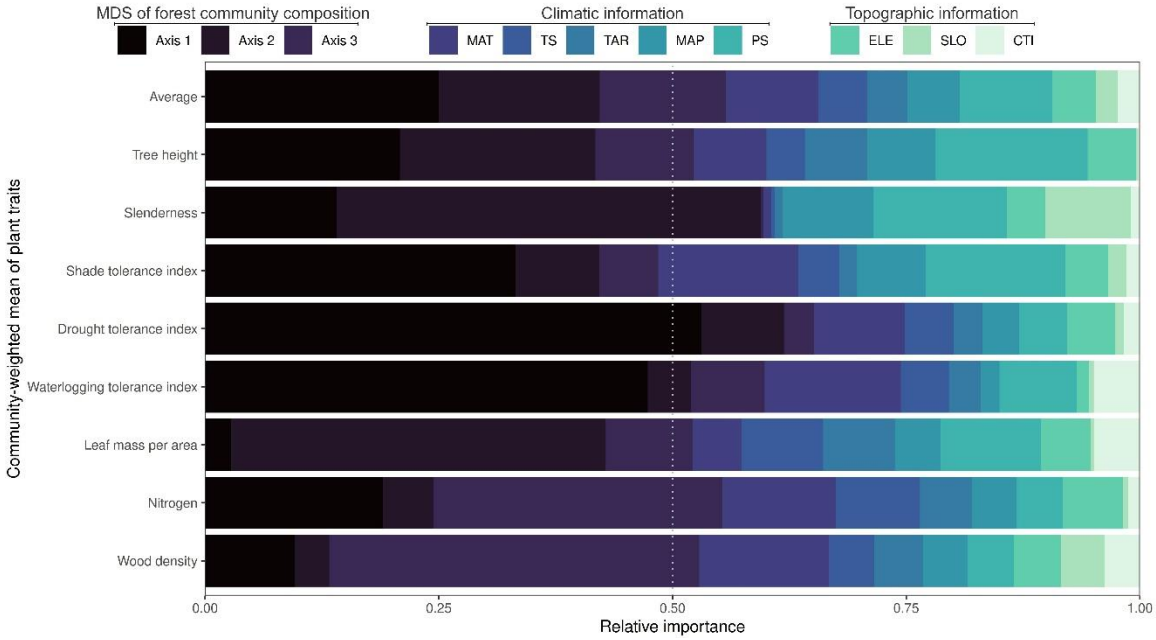


Fig S21. Relative importance of variables used in generalized linear models to predict the community-weighted mean of plant traits within inventories from the Forest Inventory and Analysis program through axes of forest composition, climatic, and topographic information. Axes (i.e., Axis 1, Axis 2, and Axis 3) describe the potential forest composition from a MDS ordination of beta diversity based on functional information. Climatic variables are described by the mean annual temperature (MAT), the temperature seasonality (TS), temperature annual range (TAR), annual precipitation (AP), and precipitation seasonality (PS). Topographic information is described by elevation (ELE), slope (SL), and compound topographic index (CTI).

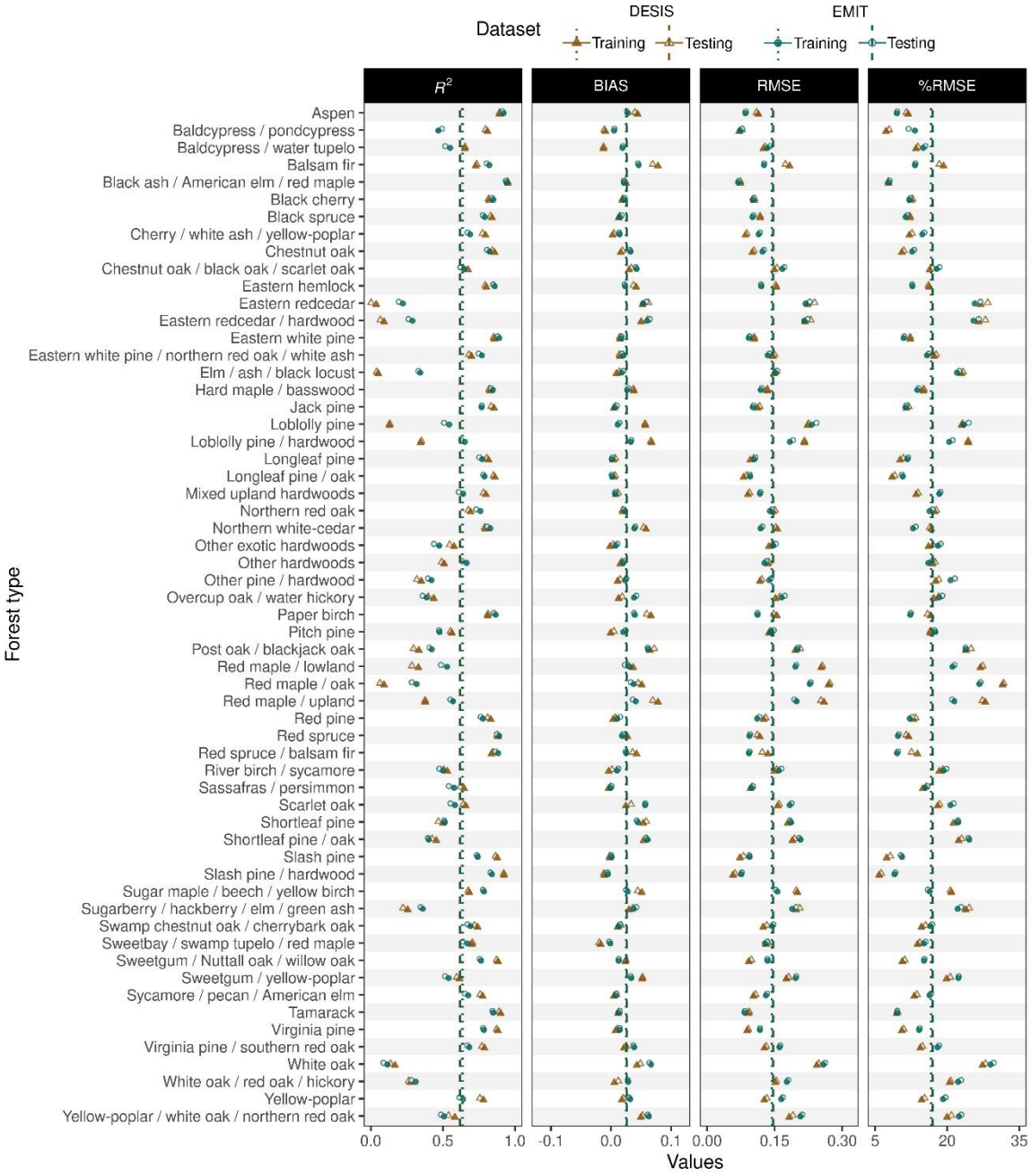


Fig S22. Performance to predict the probability of forest types of inventories within the Forest Inventory and Analysis program using climatic and topographic information as well as predicted MDS axes of dimensions of beta diversity based on spaceborne observations of DESIS or EMIT.

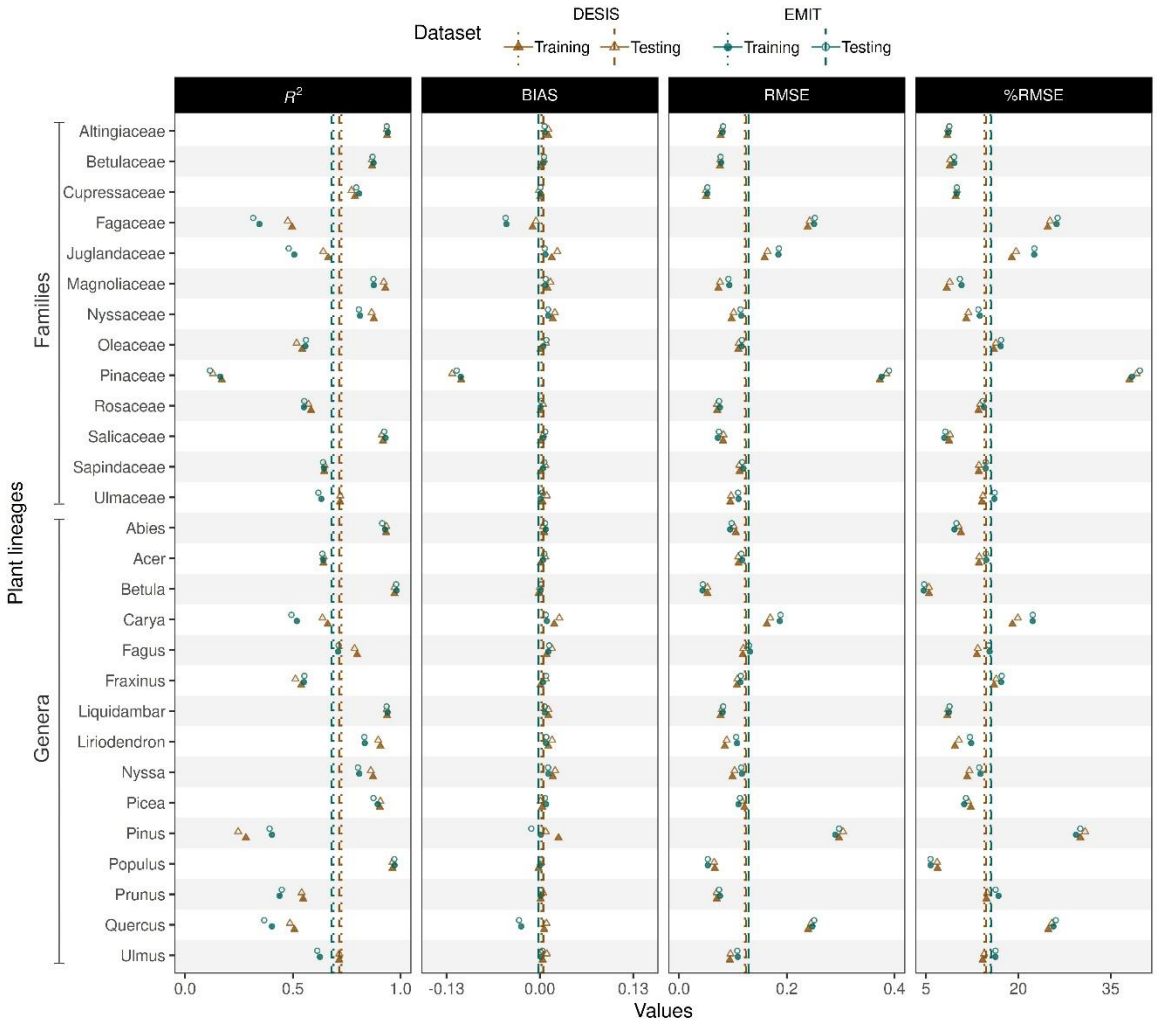


Fig S23. Performance to predict the probability of presence of plant lineages within inventories of the Forest Inventory and Analysis program using climatic and topographic information as well as predicted MDS axes of dimensions of beta diversity based on spaceborne observations of DESIS or EMIT. Each point represents the mean estimate of 100 models.

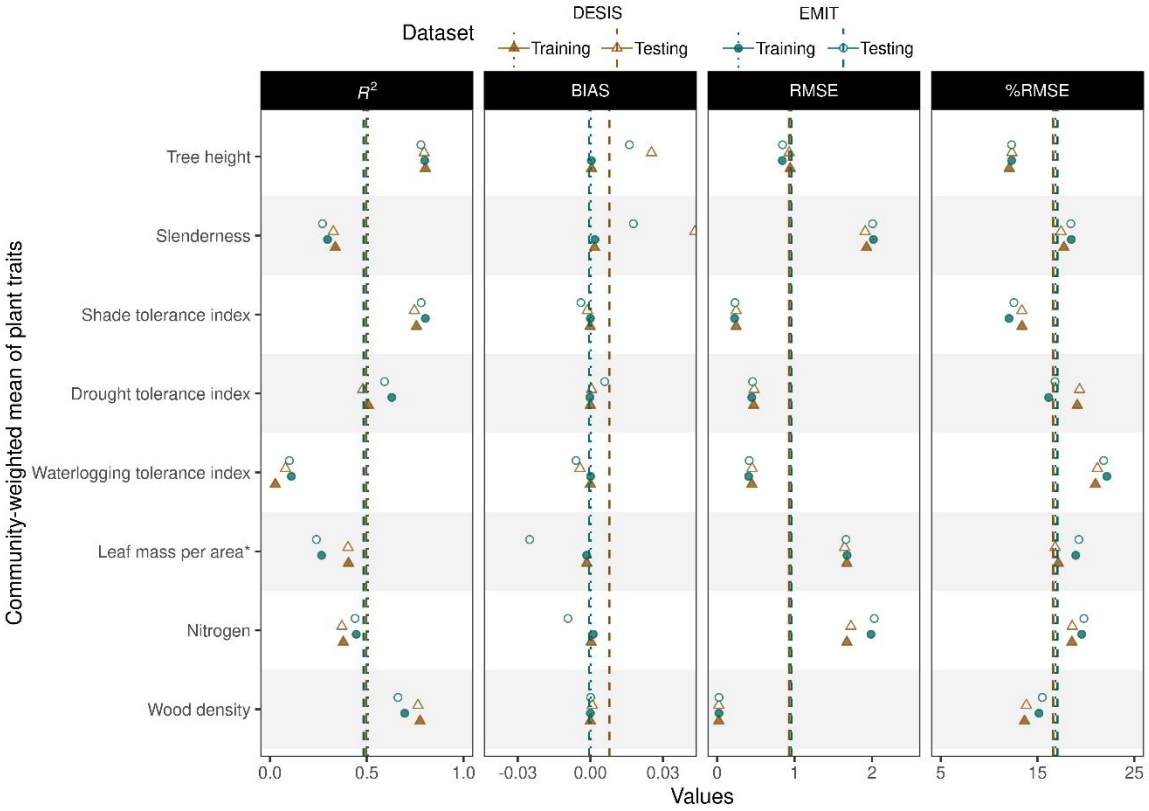


Fig S24. Performance to predict the community-weighted mean of plant traits within inventories of the Forest Inventory and Analysis program using climatic and topographic information as well as predicted MDS axes of dimensions of beta diversity based on spaceborne observations of DESIS or EMIT. Each point represents the mean estimate of 100 models.

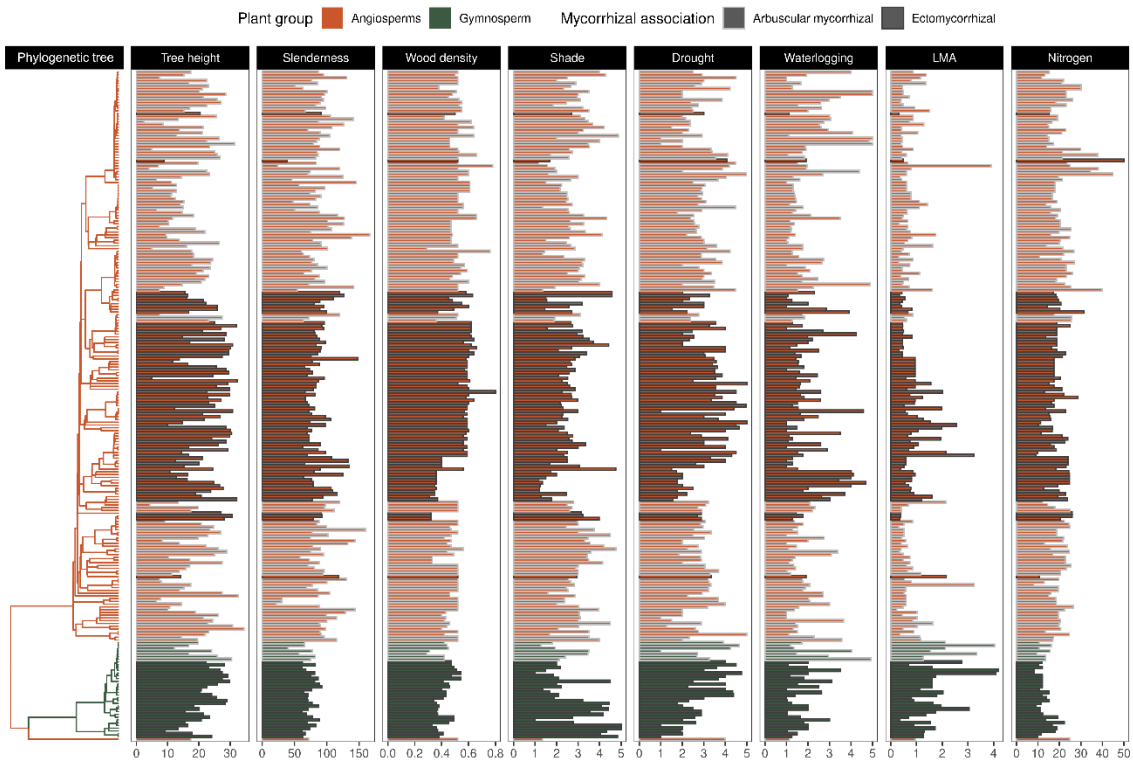


Fig S25. Plant phylogenetic tree and its functional traits of inventoried species from the Forest Inventory and Analysis program across the Eastern U.S.

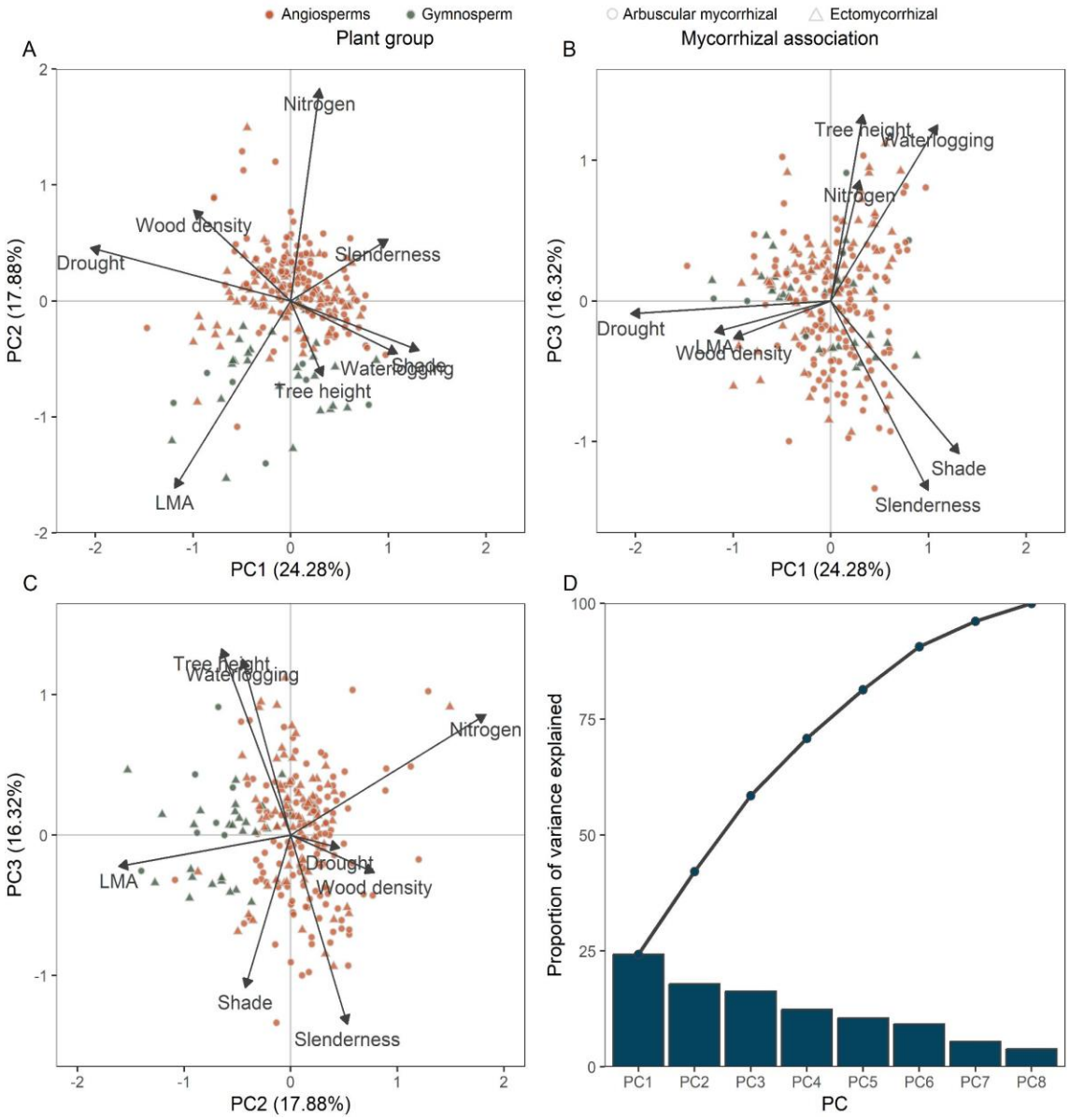


Fig S26. Principal component analysis to summarize the variability of eight plant traits among species. Panels A, B, and C describe the biplot projections among the first three principal components, while panel D describes the proportion of variance explained by each component.

SI References

1. W. W. Hargrove, F. M. Hoffman, Using multivariate clustering to characterize ecoregion borders. *Comput. Sci. Eng.* **1**, 18–25 (1999).

2. Ü. Niinemets, F. Valladares, Tolerance to shade, drought, and waterlogging of temperate Northern Hemisphere trees and shrubs. *Ecol. Monogr.* **76**, 521–547 (2006).

3. B. S. Maitner, *et al.*, The BIEN R package: A tool to access the Botanical Information and Ecology Network (BIEN) database. *Methods Ecol. Evol.* **9**, 373–379 (2018).

4. V. J. Debastiani, V. A. G. Bastazini, V. D. Pillar, Using phylogenetic information to impute missing functional trait values in ecological databases. *Ecol. Inform.* **63**, 101315 (2021).

5. S. A. Smith, J. W. Brown, Constructing a broadly inclusive seed plant phylogeny. *Am. J. Bot.* **105**, 302–314 (2018).

6. T. Santos, PVR: Phylogenetic Eigenvectors Regression and Phylogentic Signal-Representation Curve. (2018). Deposited 2018.

7. D. J. Stekhoven, P. Bühlmann, MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* **28**, 112–118 (2012).

8. E. Paradis, J. Claude, K. Strimmer, APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289–290 (2004).

9. signal developers, *signal: Signal processing* (2023).

10. S. E. Fick, R. J. Hijmans, WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017).

11. EROS, Global Topographic 30 Arc-Second Hydrologic Digital Elevation Model 1 km. U.S. Geological Survey. https://doi.org/10.5066/F77P8WN0. Deposited 2017.