Predicting species distributions and community composition using remote sensing products

Jesús N. Pinto-Ledezma^{a*} & Jeannine Cavender-Bares^a

^aDepartment of Ecology, Evolution and Behavior, University of Minnesota, 1479 Gortner Ave,
Saint Paul, MN 55108, USA

***Corresponding Author:** jpintole@umn.edu, jesuspintoledezma@gmail.com

Abstract

Accurate predictions of species composition and diversity are critical to the development of conservation actions and management strategies. In this paper using oak assemblages distributed across the conterminous United States as study model, we assessed the performance of stacked species distribution models (S-SDMs) and remote sensing products in building the next-generation of biodiversity models. This study represents the first attempt to evaluate the integrated predictions of biodiversity models—including assemblage diversity and composition—obtained by stacking next-generation SDMs. We found three main results. First, environmental predictors derived entirely from remote sensing products represent adequate covariates for biodiversity modeling. Second, applying constraints to assemblage predictions, such as imposing the probability ranking rule, results in more accurate species diversity predictions. Third, independent of the stacking procedure (bS-SDM versus cS-SDM), biodiversity models do not recover the observed species composition with high spatial resolution, i.e., correct species identities at the scale of individual plots. However, they do return reasonable predictions at macroecological scales (1 km). Our results provide insights for the prediction of assemblage diversity and composition at different spatial scales. An important task for future studies is to evaluate the reliability of combining S-SDMs with direct detection of species using image spectroscopy to build a new generation of biodiversity models to accurately predict and monitor ecological assemblages through time and space.

Keywords

Assemblage composition, biodiversity predictions, FIA, remote sensing products, species distribution models, species richness, sustainable development

1. Introduction

Species diversity and composition vary in space and time as a consequence of historical biogeographic and environmental factors and ongoing ecological processes. Since the last millennium, however, rising human population and activities have been major drivers of environmental change on Earth (Dale et al., 2000; Lewis and Maslin, 2015; Pinto-Ledezma and Rivero Mamani, 2014; Tilman et al., 2017). As a consequence, biodiversity is changing at a pace solely compared to the major extinction events recorded in the geologic history (Allen et al. 2020; IPBES, 2019; Lewis and Maslin, 2015; Pimm et al., 2006). These rapid changes in biodiversity are impacting the capacity of ecosystems to provide services to humanity that ultimately influence our well-being (Cavender-Bares et al., 2015a; Chaplin-Kramer et al., 2019; Díaz et al., 2019; Pecl et al., 2017). Thus, detecting and monitoring species diversity and composition is critical to developing effective management strategies and conservation actions facing global change (Jetz et al., 2019; Pinto-Ledezma and Rivero Mamani, 2014; Purvis et al., 2018; Watson et al., 2020) that move us towards international biodiversity goals, including those posed by the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem (IPBES) and the parallel United Nations' (UN) Sustainable Development Goals (UN-SDGs) (Díaz et al., 2019) and the upcoming post-2020 Convention on Biodiversity goals.

Different approaches have been proposed for assessing and monitoring the spatial and temporal patterns of species diversity and distribution (Jetz et al., 2019; Mateo et al., 2017; Pereira et al., 2013), including biodiversity models—i.e., models that aim to predict species diversity and composition (Ferrier and Guisan, 2006; Guisan and Rahbek, 2011) and remote sensing technologies and products (Caughlin et al., 2016; Cavender-Bares et al., 2020; Fawcett et al., 2018; Féret and Asner, 2014; Randin et al., 2020; Turner, 2014). Among biodiversity models,

the approach of Stacked Species Distribution Models (S-SDMs) has been successfully implemented for predicting species diversity and composition (D'Amen et al., 2015; Pottier et al., 2013; Zurell et al., 2020). In building S-SDMs, individual SDM are modeled as a function of environmental predictors—usually derived from interpolated climate surfaces. Subsequently, these models are stacked to produce species diversity and composition assemblage predictions (Ferrier and Guisan, 2006; Pottier et al., 2013). However, despite successful S-SDMs predictions, most previous studies were restricted to small areas (D'Amen et al., 2018, 2015; Pottier et al., 2013), limiting the potential of these models for biodiversity prediction and monitoring over large geographic areas, including at continental scales. Moreover, the uncertainty associated with the underlying data (i.e., occurrence records and environmental covariates) can negatively impact species model predictions (Dobrowski et al., 2011; Soria-Auza et al., 2010) and consequently the reliability of these kinds of biodiversity models.

Remote sensing products (RS-products) have been increasingly used to derive metrics that allow tracking biodiversity from space (Lausch et al., 2018; Rocchini et al., 2016; Schulte to Bühne and Pettorelli, 2018; Turner, 2014), monitoring the state of human impacts (Hansen et al., 2013; Pinto-Ledezma and Rivero Mamani, 2014), as predictors for describing large patterns of species diversity (Hobi et al., 2017; Radeloff et al., 2019) or to derive Essential Biodiversity Variables, i.e., measures that allow the detection and quantification of biodiversity changes (Fernández et al., 2020; Pereira et al., 2013). Despite their high spatial and temporal resolution, quasi-global coverage and range of data products (e.g., precipitation, plant productivity, biophysical variables, land cover), RS-products have been rarely used as predictors for biodiversity models (Pinto-Ledezma and Cavender-Bares, 2020; Randin et al., 2020; Saatchi et al., 2008). RS-products have been dubbed important “next-generation” environmental predictors

in biodiversity models (He et al., 2015), given that remote sensing continuously captures an increasing range of Earth's biophysical features at global scale (Cord et al., 2013; Guanter et al., 2014; Jetz et al., 2016), avoiding the uncertainty associated with environmental predictors derived from traditional climatic data. Furthermore, species models derived from RS-products perform as well as from those derived from interpolated climate surfaces have the potential to provide predictions with greater spatial resolution (Pinto-Ledezma and Cavender-Bares, 2020).

A crucial challenge for detecting and monitoring biodiversity is the integration of modeling approaches with RS-products in order to build next-generation biodiversity models at different spatial and temporal scale and levels of organization (He et al., 2015; Randin et al., 2020). Therefore, the general aim of this paper is twofold: (1) to combine S-SDMs and RS-products in order to build next-generation biodiversity models, and (2) to perform a rigorous evaluation of biodiversity model predictions. In doing so, we applied a sequential procedure that allows the prediction and evaluation of species assemblage diversity and composition at continental scale (Fig. 1). This procedure was implemented using the oak clade (genus *Quercus*) as a model study within the geographical area corresponding to the conterminous United States. We focused on this clade because oaks are widely distributed forest dominants, and they have the highest species richness and highest total biomass of any woody group in the conterminous U.S. (Cavender-Bares, 2019). Nevertheless, individual species have contrasting environmental distributions and range sizes. These attributes make the oaks both an important study system and a useful one for testing next generation SDMs.

2. Methods

2.1. Species occurrence dataset

This genus *Quercus* includes 91 species widely distributed across U.S. and showing a marked longitudinal species diversity gradient, with high species concentration in south-eastern North America. Our main occurrence dataset was assembled in a previous study (Hipp et al., 2018). We completed the main dataset using the Integrated Digitized Biocollections (iDigBio, data downloaded between 15 and 18 April, 2019) and collections from the second author (JCB) (Cavender-Bares et al., 2015b). All occurrence data were visually examined and any localities that were outside the known geographical range of the species, in unrealistic locations (e.g., water bodies, crop fields) or in botanical gardens were discarded for accuracy. In addition, to avoid problems of spatial sampling bias and spatial autocorrelation we thinned the occurrence records of each species using a spatial thinning algorithm with thinning distance of 5 km (Aiello-Lammens et al., 2015). The final dataset included 132 ± 211 (SD) unique presence data per species and 77.099% of the species had more than 25 records and 22.901% at least 10 presence records.

2.2. Environmental data derived from Remote sensing

Remote sensing products used as environmental inputs to our models included Leaf Area Index (LAI), obtained from MODIS Terra/Aqua MOD15A2 over a 15-year period (2001 – 2015) using the interface EOSDIS Earthdata (Myneni et al., 2002); precipitation from Climate Hazards group Infrared Precipitation with Stations (CHIRPS); and altitude or mean elevation from the Shuttle Radar Topography Mission (SRTM). The 15-year LAI composite provides a representation of the spatial variation of biophysical variables of different components of vegetation and ecosystems over the course of this time frame (Hobi et al., 2017; Pinto-Ledezma and Cavender-Bares, 2020). LAI strongly co-varies with the physical environment; for instance, higher LAI is

associated with warmer, wetter and stable environments, whereas lower LAI with cooler, drier and less stable environments (Gower et al., 1999; Myneni et al., 2002; Reich, 2012). In particular, LAI has a strong relationship with climate and captures the dynamics of the growing season, including vegetation seasonality—important for characterizing plant species geographical ranges (Cord et al., 2017). We posit, therefore, that the spatial and temporal variation in LAI—half of the total green leaf area per unit of horizontal ground surface area (Xiao et al., 2014)—represents an important variable for determining species distributions that integrates across biotic and abiotic factors (Peterson, 2011; Soberón, 2007). CHIRPS is a quasi-global, high resolution and daily precipitation dataset designed for monitoring drought and global environmental land change (Funk et al., 2015). SRTM is a high-resolution digital elevation model of the Earth that has been used for mapping and monitoring the earth’s surface (Farr et al., 2007) and an important variable for predicting species distributions and community composition (Dubuis et al., 2011; Pottier et al., 2013).

2.3.Species and community composition validation data

Ecological niche models represent abstractions of the environmentally suitable areas for species to maintain long-term viable populations (Peterson, 2011). The presence of a species in a locality or a grid cell informs us about the areas that are environmentally suitable for that species (Lobo et al., 2010; Peterson, 2011). Its absences inform us of the opposite pattern—namely those areas that are either not environmentally suitable for the species or are the result of historical contingences, biotic interactions that constraint the species presence even if the physical environment is suitable, recent extirpation events such as those caused by land use change, or simply because the species was not detected (Lobo et al., 2010; Peterson, 2011; Václavík and

Meentemeyer, 2009). Here, using an independent dataset of true presences and absences from the Forest Inventory and Analysis (FIA) program of the U.S. Forest Service (Smith, 2002), we evaluate the predictions of individual oak species models and also the community composition predictions. A standard FIA plot is designed to cover an area of 6052 m² (~0.6 ha); within each plot, four circular subplots of 168.3 m² were established with one center subplot and three equidistant subplots. Within each subplot every tree species with stem >2.5 cm diameter at breast height is recorded. Although the surveys are intensive, not all trees in the plot are measured because the four subplots do not cover the total area of each surveyed plot (https://www.fia.fs.fed.us/library/database-documentation/current/ver70/FIADB%20User%20Guide%20P2_7-0_ntc.final.pdf). Data of this type is commonly used in community ecology, community phylogenetics and biogeography (e.g., Cavender-Bares et al., 2018; Potter et al., 2019).

Oak species from the FIA dataset was obtained from a previous study (Cavender-Bares et al. 2018). FIA is a continuous forest census that aims to determine the extent, condition, volume, growth, and use of trees in the U.S. Shrub species are largely excluded from the FIA inventory. This dataset includes a total of 112,183 plots of 6,052 m² and 49 of the 91 oak species found in naturally assembled forests (Cavender-Bares et al., 2018). From this data set we filtered assemblages to retain only those with at least three species. The final FIA dataset used in this study includes a total of 13,263 plots and 49 oak species in the conterminous U.S.

2.4. Modeling framework

To obtain reliable SDM predictions it is necessary to define the accessible area for a species in geographical space, i.e., the area that has been accessible to the species within a given period of

time (Barve et al., 2011; Peterson, 2011). We defined the accessible area for each oak species as the combined area of the geographical ranges of sister species or the overlap of geographical ranges of a given species and its phylogenetically closest relatives (Fig. 1a). Note that the geographical ranges were defined using the occurrence records of each oak species. To do so, we first identified sister species using the published phylogeny of American oaks (Hipp et al., 2018). This phylogeny was constructed using genome-wide restriction site-association DNA sequencing (RAD-seq) and is the most highly resolved phylogenetic hypothesis for the clade in North America (Hipp et al., 2018). The rationale behind this approach is to define a large area that a species can colonize based on species environmental tolerances—i.e., under the assumption that oak species share environmental tolerances (Hipp et al., 2018; Cavender-Bares et al., 2018). Importantly, this procedure provides a conservative spatial representation of the environmental space in which a given species have potentially dispersed and detected.

Prior to model calibration, we masked the environmental variables to the extent of the defined species accessible area. This procedure was repeated for each oak species. Using an ensemble framework (Araujo and New, 2007; Diniz-Filho et al., 2009) and following Pinto-Ledezma and Cavender-Bares (2020), we modeled the ecological niche for each oak species (Fig. 1b). Under this framework, the predicted species ecological niche or distributional area is the outcome of combining the results of different modeling algorithms (Araujo and New, 2007; Diniz-Filho et al., 2009). Within this framework we fit four species models including two statistical methods (generalized linear models [GLM] and generalized additive models [GAM]) and two machine learning methods (Boosted regression trees [BRT] and Random Forest [RF]). Model evaluation was performed using a K-fold cross-validation (James et al., 2013). This approach randomly splits the data into K-folds or K-groups of relatively equal size and then fits

every model K times. At each step (K-folds) K-1 folds are used for training and remaining 1-fold for testing (James et al., 2013). We used 5 folds to split our data, i.e., 80% and 20% of the data were used for testing and training, respectively. The final ensemble model prediction for each species (Fig. 1b) was obtained using weighted averaging based on true skill statistic (Allouche et al., 2006). The ensemble procedure including model calibration, evaluation and ensemble prediction was performed using the R package sdm (Naimi and Araújo, 2016).

To obtain assemblage composition and species richness predictions we applied two stacking species distribution models (S-SDM) procedures (i.e., binary and constrained binary) (Fig. 1c). These procedures return the predicted species composition and richness within each assemblage (i.e., grid cell) across a geographical domain (Ferrier and Guisan, 2006; Guisan and Rahbek, 2011). More specifically, binary S-SDM (bS-SDM) predictions were obtained by converting individual species ensembles into binary predictions or occurrence probabilities by using its corresponding TSS-maximization threshold (Scherrer et al., 2018; Zurell et al., 2020). Constrained binary S-SDM (cS-SDM) predictions were obtained by constraining each assemblage applying a probability ranking rule (PRR). PRR emulates ecological assembly rules by ranking the species in each assemblage based on the occurrence probability obtained from each species and the number of species per assemblage. The species with the highest probabilities in an assemblage is selected until the number of species in an assemblage, based on observed data, is reached (D'Amen et al., 2015; Scherrer et al., 2018). We used the maximum number of species per assemblage from the FIA dataset as assemblage-level constraint for cS-SDM estimations.

2.5. Evaluating species and assemblage predictions

Given that the FIA dataset used in this study includes only 49 out of the 91 oak species in the U.S., we restricted the species and assemblage prediction evaluations to those 49 species (Supplementary table 1). The performance of species predictions was assessed using the Boyce index, as implemented in Hirzel et al. (2006). This index varies from -1 to +1, in which negative values indicate counter predictions and positive values indicate that model prediction are consistent with the observed distribution of presences, while values close to zero suggest that model predictions are not different from random (Boyce et al., 2002; Hirzel et al., 2006).

To evaluate the performance of species assemblage predictions, we used four different metrics that are commonly used for this purpose (Pottier et al., 2013; Scherrer et al., 2018; Zurell et al., 2020). The metrics used were: a) the deviation of the predicted species richness to the observed (SR deviation); b) the proportion of correctly predicted as present or absent (Prediction success); c) true skill statistic (TSS) and d) similarity between the observed and predicted community composition (Sørensen index). Assemblage metric evaluations were performed using the R package ecospat (Di Cola et al., 2017) using the matrices of the two assemblage predictions, i.e., binary and PRR (Fig. 1c) and 1000 iterations. Note that the SR deviation was estimated only for the bS-SDM predictions given that we used the FIA dataset in constraining the number of species per assemblage in the cS-SDM construction.

We also investigated the phylogenetic structure of oak assemblages for both the FIA dataset and the predicted assemblages, in order to explore the performance of predicted oak assemblages in recovering similar patterns of phylogenetic structure as the observed assemblages. We defined phylogenetic structure as the mean phylogenetic distance (MPD) (Webb et al., 2002). To facilitate comparison between the two datasets, we summarized the results using standardized effects sizes (SES), which compare the observed value of an

assemblage (MPD) to the mean expected value under a null model, correcting for their standard deviation. SES values >0 and <0 indicate phylogenetic clustering and overdispersion, respectively (as in Webb et al., 2002). In SES calculations we randomized the tips of the phylogeny to generate random communities (taxon shuffle null model). All phylogenetic structure calculations were conducted using customized scripts and core functions from the PhyloMeasures (Tsirogianis and Sandel, 2016) package in R.

Given the mismatch between the grain size of the observed assemblages (6,052 m² or ~ 0.006 km) and the grain size of the predictions (1 km), we were also interested in exploring the effect of increasing grain size in assemblage predictions. Accordingly, we aggregated the observed FIA assemblages into larger assemblages, ranging from 30 arcseconds (~ 1 km), 2.5 degrees (~ 25 km) and 0.5 degrees (~ 55 km), and repeated all the assemblage level analyses described above. Data aggregation was performed using the R package letsR (Vilela and Villalobos, 2015).

2.6. Statistical analysis

Using a Bayesian counterpart of Pearson's correlation test, we evaluated the relationship between species richness and phylogenetic assemblage structure obtained from both FIA and predicted datasets. We chose this Bayesian alternative because it allow robust parameter estimations and account for outliers in the data. We also used Bayesian-ANOVAs to test for differences in species assemblage predictions between different modelling procedures (i.e., stacking procedures), using plots as a random variable to correct for potential repeated measures. Using Maximum A Posteriori (MAP) p-values (Mills and Parent, 2014), we then evaluated the evidence for those differences. Note that all analyses were performed for each grain separately.

Bayesian-ANOVAs was performed using the R package rstanarm (Goodrich et al., 2020), and the Bayesian Pearson's correlation test was implemented using the R package BayesianFirstAid (https://github.com/rasmusab/bayesian_first_aid). All analyses were performed using 4 sampling chains for 10,000 generations, discarding 20% of each run as burnins. MAP-based p-values was estimated as implemented in the R package bayestestR (Makowski et al., 2019).

3. Results

3.1. Species predictions

Oak species predictions were successfully calibrated for all species (Fig. S1), showing mid to high prediction accuracy as measured by the AUC (area under the receiver operating characteristic curve) and the True skill statistic (TSS) metrics (Ensemble mean $\pm$ SD: AUC 0.858 $\pm$ 0.056, TSS 0.703 $\pm$ 0.049; Fig. S1). The performance of species predictions also showed high accuracy: all species presence values were greater than 0.5 as measured by the Boyce index (mean $\pm$ SD: 0.897 $\pm$ 0.140; Table S1), indicating that ensemble model predictions are consistent with the independent observed presences from the FIA dataset.

3.2. Assemblage predictions

Mean observed richness per assemblage at the observed grain size (i.e., FIA plots of 6052 m²) was 3.430 (SD = 0.922, max = 10, min = 3), and mean richness prediction per assemblage was 6.104 (SD = 2.443, max = 12, min = 1). Mean observed richness at 30 arc seconds (i.e., spatial resolution of the predictors) was 4.039 (SD = 1.299, max = 12, min = 3), while the prediction for richness was 8.322 (SD = 3.474, max = 16, min = 3). Bayesian correlations revealed low but

positive correlations between the observed and the predicted richness derived from the bS-SDM at the observed grain size ($\rho = 0.26$ [0.24:0.27, 95% confidence interval (CI)], Fig. 2a). Interestingly, the relationship between the observed and predicted richness increased dramatically as the grain size increased (Fig. 2b-d).

Likewise, mean absolute deviation between observed species richness and species richness predictions derived from the bS-SDM was 0.246 (SD = 0.204, max = 0.939, min = 0.000) and decreased as grain size increased (Fig. 3a). The mean assemblage prediction success for the bS-SDM was 0.885 (SD = 0.047, max = 0.983, min = 0.744; Fig. 3b) and for the cS-SDM 0.941 (SD = 0.037, max = 1, min = 0.739; Fig. 3b). Average assemblage TSS over all localities for bS-SDM were 0.436 (SD = 0.140, max = 1, min = 0.021; Fig. 3c) and 0.562 (SD = 0.276, max = 1, min = 0.121; Fig. 3c) for cS-SDM. Mean assemblage similarity estimated using the Sørensen index was 0.551 (SD = 0.125, max = 0.905, min = 0.1; Fig. 3d) for bS-SDM and 0.594 (SD = 0.258, Max = 1, min = 0.01) for cS-SDM. As expected, the accuracy of assemblage predictions increased with increasing grain size (Fig. 3). In addition, although the mean values of the metrics used to evaluate the accuracy in assemblage predictions were relatively similar between the two stacking procedures (bS-SDM vs cS-SDM), Bayesian-ANOVAs revealed strong evidence in favor for cS-SDM in predicting assemblage composition, in other words, cS-SDM stacking procedure outperformed bS-SDM. The higher performance of the cS-SDM procedure is consistent across all grain sizes, except for TSS at the grain size 0.5 (Fig. 3c), which can be considered anecdotal.

We were also interested in evaluating whether species composition of the assemblage predictions returned similar patterns of phylogenetic structure as the observed assemblages. Accordingly, the average SES-MPD at the observed grain size for the FIA dataset was -0.471

(SD = 1.130), for the bS-SDM -1.915 (SD = 1.396) and -0.201 (SD = 0.906) for the cS-SDM. In addition, most of the predicted assemblages presented negative SES-MPD values (Fig. 4), indicating that the dominant pattern of phylogenetic structure in the predicted oak assemblages is overdispersion, a pattern that is observed in natural oak assemblages. Bayesian correlations showed low but positive associations between the phylogenetic structure obtained from the observed and both assemblage predictions ($\rho = 0.15$ [0.14:0.17] and $\rho = 0.26$ [0.24:0.27], median rho and 95% CI estimated from FIA vs bS-SDM and FIA vs cS-SDM, respectively) at the observed grain size (Fig. 5a-b). Similar to the richness predictions, the correlation between the phylogenetic structure among datasets increased sharply with increasing grain size (Fig. 5d-l).

4. Discussion

The use of RS-products in biodiversity models has been hailed as a transformative approach for providing simple and flexible biodiversity predictions (He et al., 2015; Randin et al. 2020). In this study we evaluated the reliability of S-SDM—a biodiversity model (Ferrier and Guisan, 2006)—based on individual SDMs derived from RS-products only, to predict biodiversity in terms of species richness and composition within local assemblages. Our results indicate that the combination of S-SDM and RS-products perform well in predicting plot-level biodiversity as assessed by several important metrics. However, when we compared the S-SDM predictions of phylogenetic structure of oak assemblages at the plot-level with the observed structure (measured in terms of species richness and mean phylogenetic distance), the predictions tended to overestimate the diversity, even though the predicted and observed metrics were positively correlated. This result is consistent with the stacking procedure. We also found that when oak assemblages were aggregated at large grain sizes, S-SDM recovered the observed pattern well,

i.e., patterns of predicted diversity were highly similar to those obtained from the aggregated FIA plots (Fig. 2c-d, 5g-l). Consequently, our results show that biodiversity models in combination with RS-products do not necessarily predict biodiversity well at the plot-level scale, but they do well at larger grain sizes. In other words, the S-SDMs that use RS-products as environmental predictors are very useful for detecting macroecological patterns of biodiversity. These results spur new insights for biodiversity prediction and monitoring at different spatial scales.

Biodiversity models, such as the S-SDMs, depend on the reliability of individual species models (SDM) (Guisan and Rahbek, 2011). SDMs represent abstractions of species Hutchinsonian niches, i.e., species distributions are constrained by both, the abiotic environment and biotic interactions with other species (Hutchinson's duality *sensu* Colwell and Rangel, 2009; see also, Peterson et al., 2011; Pinto-Ledezma and Cavender-Bares, 2020). Our results show that individual oak SDMs constructed using environmental covariates derived from RS-products, have good accuracy (Fig. S1). However, by stacking a suite of SDMs to obtain assemblage composition predictions, no biotic constraints are considered. Consequently, S-SDM predictions tend to overestimate the number of species within local assemblages (Guisan and Rahbek, 2011; Mateo et al., 2017). We found that species richness predictions using S-SDM overpredict the number of oak species that are observed in naturally assembled oak communities (Fig. 3a). This might be related with the proportion of common (high prevalence) and rare (low prevalence) species that are co-occurring in these assemblages (Fig. S2), that in turn is affected by microhabitat variables that allow oak species to differentiate locally. Indeed, soil moisture (measured in 20x50m plots on the ground) explain oak species trait distributions better than climate variables (Cavender-Bares et al. 2020).

In addition, it has been suggested that incorporating some ecological assembly rules, e.g., the probability ranking rule approach (PRR), in S-SDM predictions can minimize assemblage overpredictions (Calabrese et al., 2014; D'Amen et al., 2015; Guisan and Rahbek, 2011). Our results corroborate this assumption; specifically, predicting species assemblages by simply overlaying individual SDMs, tends to overpredict the number of species present in local assemblages, especially at finer grains (Fig. 2a-b, 3a). Moreover, through the implementation of PRR in S-SDM predictions—i.e., by constraining the number of species present within local assemblages (Guisan and Rahbek, 2011)—we found that cS-SDM outperform bS-SDM assemblage predictions across all metrics used and grain sizes evaluated (Fig. 3b-d).

While applying constraints to assemblage predictions may be advantageous for predicting species richness (Guisan and Rahbek, 2011; Mateo et al., 2017), this is not the case for predicting species identities or composition within assemblages, particularly at finer grains. Indeed, we found that neither bS-SDM and cS-SDM accurately recover species identities at the grain size of plots (Fig. 5a-b) or at the grain size of the predictors variables (Fig. 5d-e). Not surprisingly, predictions of assemblage richness and composition improved dramatically with increasing grain size (Fig. 2c-d and 5g-k), a pattern observed independently of the implementation of PRR (cS-SDM, Fig. 5h and 5k) or its non-implementation (bS-SDM, Fig. 5g and 5j). These contrasting results might be a consequence of the fact that S-SDMs were primarily designed to assess macroecological patterns of species diversity (Ferrier and Guisan, 2006; Mateo et al., 2017). In addition, although RS-products are useful covariates for spatial modelling of biodiversity, they also reduce the uncertainty associated with environmental covariates in model predictions (Pinto-Ledezma and Cavender-Bares, 2020); consequently, the coarse grain of current RS-products hampers accurate predictions of biodiversity at ecological scales (Gamon et al., 2020; Jetz et al.,

2016, Pinto-Ledezma and Cavender-Bares, 2020), in other words, the broad spatial resolution of current RS-products (usually with a pixel size of 1 km) limits our ability to capture environmental features at ecological scales (e.g., microtopography, soil moisture, landscape structure) that ultimately determine the coexistence of species locally. Furthermore, we acknowledge that we evaluated only one type of biodiversity model—S-SDM using two stacking procedure—and that other biodiversity models such as the Joint-SDM (Pollock et al., 2014), could potentially improve biodiversity predictions. Nevertheless, a recent study suggests that Joint-SDM do not improve species assemblage predictions (Zurell et al., 2020). Further research is needed to correctly predict biodiversity at the assemblage level.

The evaluation presented here is meant to stimulate further methodological and empirical research to better predict biodiversity at different spatial and temporal scales and levels of organization. A promising approach for this purpose is the hierarchical modeling of species distributions (Mateo et al., 2019; Petitpierre et al., 2016). H-SDMs allow the simultaneous modeling spatial patterns of biodiversity at ecological and regional scales. In constructing HSDMs, individual SDMs are first fit at regional and landscape scales, and the two predictions are fused to obtain a single species SDM, i.e., the regional model is rescaled to the landscape model (Mateo et al., 2019; Petitpierre et al., 2016). This is an interesting approach because it fuses the benefits of macroecological covariates (derived from climatic surfaces or RS-products) with those derived from high-resolution remote sensing products (Mateo et al., 2019).

An unparalleled alternative is the direct detection of species using imaging spectroscopy (Cavender-Bares et al., 2017; Schweiger et al., 2020). For example, leaf-spectra variation among individual plants obtained using imaging spectroscopy provide sufficient information for the correct assignment of populations to species to clades (Cavender-Bares et al., 2016; Meireles et

al., 2020; Schweiger et al., 2020) and airborne imagery accurately assigns vegetation canopies to species (Foster and Townsend, 2004). Current and forthcoming hyperspectral images from DESIS sensors and forthcoming SBG and CHIMES sensors, among others (Alonso et al., 2019; Stavros et al., 2017; Turner, 2014) will capture information from the Earth at fine spectral resolution, allowing the estimation of plant traits, plant nutrient content, biophysical variables (e.g., leaf area index, biomass), that can be used for direct detection of functional and perhaps community diversity from space (Jetz et al., 2016). Combining the detection of species using imaging spectroscopy with SDMs to build a new generation of biodiversity models, may open new avenues for the accurately assignment of species and assessment of ecological assemblages through space and time (Féret and Asner, 2014; Meireles et al., 2020; Randin et al., 2020), a critical hurdle to overcome in addressing the challenges posed by the global change.

5. Conclusion

Recent review papers (He et al., 2015; Randin et al., 2020) proposed the applicability of remote sensing data as environmental covariates in the construction of next-generation SDMs. Here using environmental covariates derived completely from remote sensing data, we modelled the distribution of oaks (genus *Quercus*) and predicted the number and composition of species within assemblages at different grain sizes. Despite high variability in the predictions, modeled oak assemblages showed phylogenetic overdispersion, indicating that models recovered the observed pattern of distantly related oak species co-occurring more often than expected. Overall we conclude that species richness can be predicted with high accuracy by applying constraints to the predictions, however, accurate predictions of species identities is still an evolving task. We suggest two alternatives (i.e., H-SDMs and the direct detection of species using imaging

spectroscopy) that might increase the accuracy in assemblage composition predictions, hence, future studies should evaluate the reliability of these two alternatives using different taxa and across geographical settings.

Data Accessibility

All data used in this paper are already published or publicly available.

Author contribution statement

JNP-L and JC-B conceived the ideas presented and tested herein and contributed throughout the whole writing process.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

J.N.P-L. was supported by the University of Minnesota College of Biological Sciences' Grand Challenges in Biology Postdoctoral Program.

References

- Aiello-Lammens, M.E., Boria, R.A., Radosavljevic, A., Vilela, B., Anderson, R.P., 2015. spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography* 38, 541–545. <https://doi.org/10.1111/ecog.01132>
- Allen, B.J., Wignall, P.B., Hill, D.J., Saupe, E.E., Dunhill, A.M. 2020. The latitudinal diversity gradient of tetrapods across the Permo-Triassic mass extinction and recovery interval. *Proc. R. Soc. B.* 287(1929), 20201125. doi:10.1098/rspb.2020.1125
- Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS): Assessing the accuracy of distribution models. *J. Appl. Ecol.* 43, 1223–1232. <https://doi.org/10.1111/j.1365-2664.2006.01214.x>
- Alonso, K., Bachmann, M., Burch, K., Carmona, E., Cerra, D., de los Reyes, R., Dietrich, D., Heiden, U., Hölderlin, A., Ickes, J., Knodt, U., Krutz, D., Lester, H., Müller, R., Pagnutti, M., Reinartz, P., Richter, R., Ryan, R., Sebastian, I., Tegler, M., 2019. Data Products, Quality and Validation of the DLR Earth Sensing Imaging Spectrometer (DESI). *Sensors* 19, 4471. <https://doi.org/10.3390/s19204471>
- Araujo, M., New, M., 2007. Ensemble forecasting of species distributions. *Trends Ecol. Evol.* 22, 42–47. <https://doi.org/10.1016/j.tree.2006.09.010>
- Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S.P., Peterson, A.T., Soberón, J., Villalobos, F., 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecol. Model.* 222, 1810–1819. <https://doi.org/10.1016/j.ecolmodel.2011.02.011>

475 Boyce, M.S., Vernier, P.R., Nielsen, S.E., Schmiegelow, F.K.A., 2002. Evaluating resource
476 selection functions. *Ecol. Model.* 157, 281–300. <https://doi.org/10.1016/S0304->
477 3800(02)00200-4

478 Calabrese, J.M., Certain, G., Kraan, C., Dormann, C.F., 2014. Stacking species distribution
479 models and adjusting bias by linking them to macroecological models: Stacking species
480 distribution models. *Glob. Ecol. Biogeogr.* 23, 99–112. <https://doi.org/10.1111/geb.12102>

481 Caughlin, T.T., Graves, S.J., Asner, G.P., van Breugel, M., Hall, J., Martin, R.E., Ashton, M.S.,
482 Bohlman, S.A., 2016. A hyperspectral image can predict tropical tree growth rates in
483 single-species stands. *Ecol. Appl.* 26, 2369–2375. <https://doi.org/10.1002/eap.1436>

484 Cavender-Bares, J., 2019. Diversification, adaptation, and community assembly of the American
485 oaks (*Quercus*), a model clade for integrating ecology and evolution. *New Phytol.* 221,
486 669–692. <https://doi.org/10.1111/nph.15450>

487 Cavender-Bares, J., Balvanera, P., King, E., Polasky, S., 2015a. Ecosystem service trade-offs
488 across global contexts and scales. *Ecol. Soc.* 20, art22. <https://doi.org/10.5751/ES-07137->
489 200122

490 Cavender-Bares, J., Gamon, J.A., Hobbie, S.E., Madritch, M.D., Meireles, J.E., Schweiger, A.K.,
491 Townsend, P.A., 2017. Harnessing plant spectra to integrate the biodiversity sciences
492 across biological and spatial scales. *Am. J. Bot.* 104, 966–969.
493 <https://doi.org/10.3732/ajb.1700061>

494 Cavender-Bares, J., Gamon, J.A., Townsend, P.A., 2020. Remote Sensing of Plant Biodiversity.
495 Springer International Publishing.

496 Cavender-Bares, J., González-Rodríguez, A., Eaton, D.A.R., Hipp, A.A.L., Beulke, A., Manos,
497 P.S., 2015b. Phylogeny and biogeography of the American live oaks (*Quercus*

subsection *Virentes*): a genomic and population genetics approach. Mol. Ecol. 24, 3668–3687. <https://doi.org/10.1111/mec.13269>

Cavender-Bares, J., Kothari, S., Meireles, J.E., Kaproth, M.A., Manos, P.S., Hipp, A.L., 2018. The role of diversification in community assembly of the oaks (*Quercus* L.) across the continental U.S. Am. J. Bot. 105, 565–586. <https://doi.org/10.1002/ajb2.1049>

Cavender-Bares, J., Meireles, J., Couture, J., Kaproth, M., Kingdon, C., Singh, A., Serbin, S., Center, A., Zuniga, E., Pilz, G., Townsend, P., 2016. Associations of Leaf Spectra with Genetic and Phylogenetic Variation in Oaks: Prospects for Remote Detection of Biodiversity. Remote Sens. 8, 221. <https://doi.org/10.3390/rs8030221>

Cavender-Bares, J., G. Fontes, C., Pinto-Ledezma, J.N., 2020. Open questions in understanding the adaptive significance of plant functional trait variation within a single lineage. New Phytol. 227, 659–663. <https://doi.org/10.1111/nph.16652>

Chaplin-Kramer, R., Sharp, R.P., Weil, C., Bennett, E.M., Pascual, U., Arkema, K.K., Brauman, K.A., Bryant, B.P., Guerry, A.D., Haddad, N.M., Hamann, M., Hamel, P., Johnson, J.A., Mandle, L., Pereira, H.M., Polasky, S., Ruckelshaus, M., Shaw, M.R., Silver, J.M., Vogl, A.L., Daily, G.C., 2019. Global modeling of nature’s contributions to people. Science 366, 255–258. <https://doi.org/10.1126/science.aaw3372>

Colwell, R. K., & Rangel, T. F. (2009). Hutchinson’s duality: The once and future niche. Proc. Natl. Acad. Sci. USA. 106, 19651–19658. <https://doi.org/10.1073/pnas.0901650106>

Cord, A.F., Brauman, K.A., Chaplin-Kramer, R., Huth, A., Ziv, G., Seppelt, R., 2017. Priorities to Advance Monitoring of Ecosystem Services Using Earth Observation. Trends Ecol. Evol. 32, 416–428. <https://doi.org/10.1016/j.tree.2017.03.003>

520 Cord, A.F., Meentemeyer, R.K., Leitão, P.J., Václavík, T., 2013. Modelling species distributions
521 with remote sensing data: bridging disciplinary perspectives. *J. Biogeogr.* 40, 2226–2227.
522 <https://doi.org/10.1111/jbi.12199>

523 Dale, V.H., Brown, S., Haeuber, R.A., Hobbs, N.T., Huntly, N., Naiman, R.J., Riebsame, W.E.,
524 Turner, M.G., Valone, T.J., 2000. Ecological Principles and Guidelines for Managing the
525 Use of Land. *Ecol. Appl.* 10, 639. <https://doi.org/10.2307/2641032>

526 D’Amen, M., Mateo, R.G., Pottier, J., Thuiller, W., Maiorano, L., Pellissier, L., Ndiribe, C.,
527 Salamin, N., Guisan, A., 2018. Improving spatial predictions of taxonomic, functional
528 and phylogenetic diversity. *J. Ecol.* 106, 76–86. <https://doi.org/10.1111/1365-2745.12801>

529 D’Amen, M., Pradervand, J.-N., Guisan, A., 2015. Predicting richness and composition in
530 mountain insect communities at high resolution: a new test of the SESAM framework:
531 Community-level models of insects. *Glob. Ecol. Biogeogr.* 24, 1443–1453.
532 <https://doi.org/10.1111/geb.12357>

533 Di Cola, V., Broennimann, O., Petitpierre, B., Breiner, F.T., D’Amen, M., Randin, C., Engler,
534 R., Pottier, J., Pio, D., Dubuis, A., Pellissier, L., Mateo, R.G., Hordijk, W., Salamin, N.,
535 Guisan, A., 2017. ecospat: an R package to support spatial analyses and modeling of
536 species niches and distributions. *Ecography* 40, 774–787.
537 <https://doi.org/10.1111/ecog.02671>

538 Díaz, S., Settele, J., Brondízio, E.S., Ngo, H.T., Agard, J., Arneth, A., Balvanera, P., Brauman,
539 K.A., Butchart, S.H.M., Chan, K.M.A., Garibaldi, L.A., Ichii, K., Liu, J., Subramanian,
540 S.M., Midgley, G.F., Miloslavich, P., Molnár, Z., Obura, D., Pfaff, A., Polasky, S.,
541 Purvis, A., Razzaque, J., Reyers, B., Chowdhury, R.R., Shin, Y.-J., Visseren-Hamakers,
542 I., Willis, K.J., Zayas, C.N., 2019. Pervasive human-driven decline of life on Earth points

to the need for transformative change. *Science* 366, eaax3100.

<https://doi.org/10.1126/science.aax3100>

Diniz-Filho, J.A.F., Mauricio Bini, L., Fernando Rangel, T., Loyola, R.D., Hof, C., Nogués-Bravo, D., Araújo, M.B., 2009. Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. *Ecography* 32, 897–906.

<https://doi.org/10.1111/j.1600-0587.2009.06196.x>

Dobrowski, S.Z., Thorne, J.H., Greenberg, J.A., Safford, H.D., Mynsberge, A.R., Crimmins, S.M., Swanson, A.K., 2011. Modeling plant ranges over 75 years of climate change in California, USA: temporal transferability and species traits. *Ecol. Monogr.* 81, 241–257.

<https://doi.org/10.1890/10-1325.1>

Dubuis, A., Pottier, J., Rion, V., Pellissier, L., Theurillat, J.-P., Guisan, A., 2011. Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking modelling approaches: Predicting plant species richness. *Divers. Distrib.* 17, 1122–1131. <https://doi.org/10.1111/j.1472-4642.2011.00792.x>

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D., 2007. The Shuttle Radar Topography Mission. *Rev. Geophys.* 45, RG2004. <https://doi.org/10.1029/2005RG000183>

Fawcett, D., Verhoef, W., Schlöpfer, D., Schneider, F.D., Schaepman, M.E., Damm, A., 2018. Advancing retrievals of surface reflectance and vegetation indices over forest ecosystems by combining imaging spectroscopy, digital object models, and 3D canopy modelling. *Remote Sens. Environ.* 204, 583–595. <https://doi.org/10.1016/j.rse.2017.09.040>

Féret, J.-B., Asner, G.P., 2014. Mapping tropical forest canopy diversity using high-fidelity imaging spectroscopy. *Ecol. Appl.* 24, 1289–1296. <https://doi.org/10.1890/13-1824.1>

Fernández, N., Ferrier, S., Navarro, L.M., Pereira, H.M., 2020. Essential Biodiversity Variables: Integrating In-Situ Observations and Remote Sensing Through Modeling, in: Cavender-Bares, J., Gamon, J.A., Townsend, P.A. (Eds.), *Remote Sensing of Plant Biodiversity*. Springer International Publishing, Cham, pp. 485–501. https://doi.org/10.1007/978-3-030-33157-3_18

Ferrier, S., Guisan, A., 2006. Spatial modelling of biodiversity at the community level. *J. Appl. Ecol.* 43, 393–404. <https://doi.org/10.1111/j.1365-2664.2006.01149.x>

Foster, J.R., P. A. Townsend, P.A., 2004. Linking hyperspectral imagery and forest inventories for forest assessment in the Central Appalachians. Pages 76-86. U.S. Department of Agriculture, Forest Service, Northeastern Research Station.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., Michaelsen, J., 2015. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Sci. Data* 2, 150066. <https://doi.org/10.1038/sdata.2015.66>

Gamon, J.A., Wang, R., Gholizadeh, H., Zutta, B., Townsend, P.A., Cavender-Bares, J., 2020. Consideration of Scale in Remote Sensing of Biodiversity, in: Cavender-Bares, J., Gamon, J.A., Townsend, Philip A. (Eds.), *Remote Sensing of Plant Biodiversity*. Springer International Publishing, Cham, pp. 425–447. https://doi.org/10.1007/978-3-030-33157-3_16

586 Gower, S.T., Kucharik, C.J., Norman, J.M., 1999. Direct and Indirect Estimation of Leaf Area
587 Index, fAPAR, and Net Primary Production of Terrestrial Ecosystems. *Remote Sens.*
588 *Environ.* 70, 29–51. [https://doi.org/10.1016/S0034-4257\(99\)00056-5](https://doi.org/10.1016/S0034-4257(99)00056-5)

589 Guanter, L., Zhang, Y., Jung, M., Joiner, J., Voigt, M., Berry, J.A., Frankenberg, C., Huete,
590 A.R., Zarco-Tejada, P., Lee, J.-E., Moran, M.S., Ponce-Campos, G., Beer, C., Camps-
591 Valls, G., Buchmann, N., Gianelle, D., Klumpp, K., Cescatti, A., Baker, J.M., Griffis,
592 T.J., 2014. Global and time-resolved monitoring of crop photosynthesis with chlorophyll
593 fluorescence. *Proc. Natl. Acad. Sci.* 111, E1327–E1333.
594 <https://doi.org/10.1073/pnas.1320008111>

595 Guisan, A., Rahbek, C., 2011. SESAM - a new framework integrating macroecological and
596 species distribution models for predicting spatio-temporal patterns of species
597 assemblages: Predicting spatio-temporal patterns of species assemblages. *J. Biogeogr.* 38,
598 1433–1444. <https://doi.org/10.1111/j.1365-2699.2011.02550.x>

599 Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau,
600 D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L.,
601 Justice, C.O., Townshend, J.R.G., 2013. High-Resolution Global Maps of 21st-Century
602 Forest Cover Change. *Science* 342, 850–853. <https://doi.org/10.1126/science.1244693>

603 He, K.S., Bradley, B.A., Cord, A.F., Rocchini, D., Tuanmu, M.-N., Schmidtlein, S., Turner, W.,
604 Wegmann, M., Pettorelli, N., 2015. Will remote sensing shape the next generation of
605 species distribution models? *Remote Sens. Ecol. Conserv.* 1, 4–18.
606 <https://doi.org/10.1002/rse2.7>

607 Hipp, A.L., Manos, P.S., González-Rodríguez, A., Hahn, M., Kaproth, M., McVay, J.D., Avalos,
608 S.V., Cavender-Bares, J., 2018. Sympatric parallel diversification of major oak clades in

609 the Americas and the origins of Mexican species diversity. *New Phytol.* 217, 439–452.
610 <https://doi.org/10.1111/nph.14773>

611 Hirzel, A.H., Le Lay, G., Helfer, V., Randin, C., Guisan, A., 2006. Evaluating the ability of
612 habitat suitability models to predict species presences. *Ecol. Model.* 199, 142–152.
613 <https://doi.org/10.1016/j.ecolmodel.2006.05.017>

614 Hobi, M.L., Dubinin, M., Graham, C.H., Coops, N.C., Clayton, M.K., Pidgeon, A.M., Radeloff,
615 V.C., 2017. A comparison of Dynamic Habitat Indices derived from different MODIS
616 products as predictors of avian species richness. *Remote Sens. Environ.* 195, 142–152.
617 <https://doi.org/10.1016/j.rse.2017.04.018>

618 IPBES, 2019. Summary for policymakers of the global assessment report on biodiversity and
619 ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and
620 Ecosystem Services. IPBES secretariat, Bonn, Germany

621 James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. *An Introduction to Statistical Learning*,
622 Springer Texts in Statistics. Springer New York, New York, NY.
623 <https://doi.org/10.1007/978-1-4614-7138-7>

624 Jetz, W., Cavender-Bares, J., Pavlick, R., Schimel, D., Davis, F.W., Asner, G.P., Guralnick, R.,
625 Kattge, J., Latimer, A.M., Moorcroft, P., Schaepman, M.E., Schildhauer, M.P.,
626 Schneider, F.D., Schrod, F., Stahl, U., Ustin, S.L., 2016. Monitoring plant functional
627 diversity from space. *Nat. Plants* 2, 16024. <https://doi.org/10.1038/nplants.2016.24>

628 Jetz, W., McGeoch, M.A., Guralnick, R., Ferrier, S., Beck, J., Costello, M.J., Fernandez, M.,
629 Geller, G.N., Keil, P., Merow, C., Meyer, C., Muller-Karger, F.E., Pereira, H.M., Regan,
630 E.C., Schmeller, D.S., Turak, E., 2019. Essential biodiversity variables for mapping and

monitoring species populations. *Nat. Ecol. Evol.* 3, 539–551.

<https://doi.org/10.1038/s41559-019-0826-1>

Lausch, A., Bastian, O., Klotz, S., Leitão, P.J., Jung, A., Rocchini, D., Schaepman, M.E., Skidmore, A.K., Tischendorf, L., Knapp, S., 2018. Understanding and assessing vegetation health by in situ species and remote-sensing approaches. *Methods Ecol. Evol.* 9, 1799–1809. <https://doi.org/10.1111/2041-210X.13025>

Lewis, S.L., Maslin, M.A., 2015. Defining the Anthropocene. *Nature* 519, 171–180. <https://doi.org/10.1038/nature14258>

Lobo, J.M., Jiménez-Valverde, A., Hortal, J., 2010. The uncertain nature of absences and their importance in species distribution modelling. *Ecography* 33, 103–114. <https://doi.org/10.1111/j.1600-0587.2009.06039.x>

Makowski, D., Ben-Shachar, M., Lüdtke, D., 2019. bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. *J. Open Source Softw.* 4, 1541. <https://doi.org/10.21105/joss.01541>

Mateo, R.G., Gastón, A., Aroca-Fernández, M.J., Broennimann, O., Guisan, A., Saura, S., García-Viñas, J.I., 2019. Hierarchical species distribution models in support of vegetation conservation at the landscape scale. *J. Veg. Sci.* 30, 386–396. <https://doi.org/10.1111/jvs.12726>

Mateo, R.G., Mokany, K., Guisan, A., 2017. Biodiversity Models: What If Unsaturation Is the Rule? *Trends Ecol. Evol.* 32, 556–566. <https://doi.org/10.1016/j.tree.2017.05.003>

Meireles, J.E., O’Meara, B., Cavender-Bares, J., 2020. Linking Leaf Spectra to the Plant Tree of Life, in: Cavender-Bares, J., Gamon, J.A., Townsend, P.A. (Eds.), *Remote Sensing of*

653 Plant Biodiversity. Springer International Publishing, Cham, pp. 155–172.

654 https://doi.org/10.1007/978-3-030-33157-3_7

655 Mills, J.A., Parent, O., 2014. Bayesian MCMC Estimation, in: Fischer, M.M., Nijkamp, P.

656 (Eds.), Handbook of Regional Science. Springer Berlin Heidelberg, Berlin, Heidelberg,

657 pp. 1571–1595. https://doi.org/10.1007/978-3-642-23430-9_89

658 Myneni, R.B., Hoffman, S., Knyazikhin, Y., Privette, J.L., Glassy, J., Tian, Y., Wang, Y., Song,

659 X., Zhang, Y., Smith, G.R., Lotsch, A., Friedl, M., Morisette, J.T., Votava, P., Nemani,

660 R.R., Running, S.W., 2002. Global products of vegetation leaf area and fraction absorbed

661 PAR from year one of MODIS data. Remote Sens. Environ. 83, 214–231.

662 [https://doi.org/10.1016/S0034-4257\(02\)00074-3](https://doi.org/10.1016/S0034-4257(02)00074-3)

663 Naimi, B., Araújo, M.B., 2016. sdm: a reproducible and extensible R platform for species

664 distribution modelling. Ecography 39, 368–375. <https://doi.org/10.1111/ecog.01881>

665 Pecl, G.T., Araújo, M.B., Bell, J.D., Blanchard, J., Bonebrake, T.C., Chen, I.-C., Clark, T.D.,

666 Colwell, R.K., Danielsen, F., Evengård, B., Falconi, L., Ferrier, S., Frusher, S., Garcia,

667 R.A., Griffis, R.B., Hobday, A.J., Janion-Scheepers, C., Jarzyna, M.A., Jennings, S.,

668 Lenoir, J., Linnetved, H.I., Martin, V.Y., McCormack, P.C., McDonald, J., Mitchell, N.J.,

669 Mustonen, T., Pandolfi, J.M., Pettorelli, N., Popova, E., Robinson, S.A., Scheffers, B.R.,

670 Shaw, J.D., Sorte, C.J.B., Strugnell, J.M., Sunday, J.M., Tuanmu, M.-N., Vergés, A.,

671 Villanueva, C., Wernberg, T., Wapstra, E., Williams, S.E., 2017. Biodiversity

672 redistribution under climate change: Impacts on ecosystems and human well-being.

673 Science 355, eaai9214. <https://doi.org/10.1126/science.aai9214>

674 Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J., Bruford,

675 M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C., Coops, N.C., Dulloo, E., Faith,

676 D.P., Freyhof, J., Gregory, R.D., Heip, C., Hoft, R., Hurtt, G., Jetz, W., Karp, D.S.,
677 McGeoch, M.A., Obura, D., Onoda, Y., Pettorelli, N., Reyers, B., Sayre, R.,
678 Scharlemann, J.P.W., Stuart, S.N., Turak, E., Walpole, M., Wegmann, M., 2013.
679 Essential Biodiversity Variables. *Science* 339, 277–278.
680 <https://doi.org/10.1126/science.1229931>

681 Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura,
682 M., Araújo, M., 2011. *Ecological Niches and Geographic Distributions* (MPB-49).
683 Princeton University Press, Princeton, N.J. doi:10.2307/j.ctt7stnh.

684 Petitpierre, B., McDougall, K., Seipel, T., Broennimann, O., Guisan, A., Kueffer, C., 2016. Will
685 climate change increase the risk of plant invasions into mountains? *Ecol. Appl.* 26, 530–
686 544. <https://doi.org/10.1890/14-1871>

687 Pimm, S., Raven, P., Peterson, A., Sekercioglu, C. H., Ehrlich, P. R., 2006. Human impacts on
688 the rates of recent, present, and future bird extinctions. *Proc. Natl. Acad. Sci. USA.*
689 103(29), 10941–10946. <https://doi.org/10.1073/pnas.0604181103>

690 Pinto-Ledezma, J.N., Rivero Mamani, M.L., 2014. Temporal patterns of deforestation and
691 fragmentation in lowland Bolivia: implications for climate change. *Clim. Change* 127,
692 43–54. <https://doi.org/10.1007/s10584-013-0817-1>

693 Pinto-Ledezma, J.N., Cavender-Bares, J., 2020. Using remote sensing for modeling and
694 monitoring species distributions. In: Cavender-Bares, J., Gamon, J., Townsend, P. (Eds.),
695 *Remote Sensing of Plant Biodiversity*. Springer.

696 Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O’Hara, R.B., Parris, K.M., Vesk, P.A.,
697 McCarthy, M.A., 2014. Understanding co-occurrence by modelling species

simultaneously with a Joint Species Distribution Model (JSDM). *Methods Ecol. Evol.* 5, 397–406. <https://doi.org/10.1111/2041-210X.12180>

Potter, K. M., Escanferla, M. E., Jetton, R. M., Man, G., Crane, B. S., 2019. Prioritizing the conservation needs of United States tree species: Evaluating vulnerability to forest insect and disease threats. *Glob. Ecol. Conserv.* 18, e00622. doi:10.1016/j.gecco.2019.e00622

Pottier, J., Dubuis, A., Pellissier, L., Maiorano, L., Rossier, L., Randin, C.F., Vittoz, P., Guisan, A., 2013. The accuracy of plant assemblage prediction from species distribution models varies along environmental gradients: Climate and species assembly predictions. *Glob. Ecol. Biogeogr.* 22, 52–63. <https://doi.org/10.1111/j.1466-8238.2012.00790.x>

Purvis, A., Newbold, T., De Palma, A., Contu, S., Hill, S.L.L., Sanchez-Ortiz, K., Phillips, H.R.P., Hudson, L.N., Lysenko, I., Börger, L., Scharlemann, J.P.W., 2018. Modelling and Projecting the Response of Local Terrestrial Biodiversity Worldwide to Land Use and Related Pressures: The PREDICTS Project, in: *Advances in Ecological Research*. Elsevier, pp. 201–241. <https://doi.org/10.1016/bs.aecr.2017.12.003>

Radeloff, V.C., Dubinin, M., Coops, N.C., Allen, A.M., Brooks, T.M., Clayton, M.K., Costa, G.C., Graham, C.H., Helmers, D.P., Ives, A.R., Kolesov, D., Pidgeon, A.M., Rapacciuolo, G., Razenkova, E., Suttidate, N., Young, B.E., Zhu, L., Hobi, M.L., 2019. The Dynamic Habitat Indices (DHIs) from MODIS and global biodiversity. *Remote Sens. Environ.* 222, 204–214. <https://doi.org/10.1016/j.rse.2018.12.009>

Randin, C.F., Ashcroft, M.B., Bolliger, J., Cavender-Bares, J., Coops, N.C., Dullinger, S., Dirnböck, T., Eckert, S., Ellis, E., Fernández, N., Giuliani, G., Guisan, A., Jetz, W., Joost, S., Karger, D., Lembrechts, J., Lenoir, J., Luoto, M., Morin, X., Price, B., Rocchini, D., Schaepman, M., Schmid, B., Verburg, P., Wilson, A., Woodcock, P., Yoccoz, N., Payne,

D., 2020. Monitoring biodiversity in the Anthropocene using remote sensing in species distribution models. *Remote Sens. Environ.* 239, 111626.
<https://doi.org/10.1016/j.rse.2019.111626>

Reich, P.B., 2012. Key canopy traits drive forest productivity. *Proc. R. Soc. B Biol. Sci.* 279, 2128–2134. <https://doi.org/10.1098/rspb.2011.2270>

Rocchini, D., Boyd, D.S., Féret, J.-B., Foody, G.M., He, K.S., Lausch, A., Nagendra, H., Wegmann, M., Pettorelli, N., 2016. Satellite remote sensing to monitor species diversity: potential and pitfalls. *Remote Sens. Ecol. Conserv.* 2, 25–36.
<https://doi.org/10.1002/rse2.9>

Saatchi, S., Buermann, W., ter Steege, H., Mori, S., Smith, T.B., 2008. Modeling distribution of Amazonian tree species and diversity using remote sensing measurements. *Remote Sens. Environ.* 112, 2000–2017. <https://doi.org/10.1016/j.rse.2008.01.008>

Scherrer, D., D’Amen, M., Fernandes, R.F., Mateo, R.G., Guisan, A., 2018. How to best threshold and validate stacked species assemblages? Community optimisation might hold the answer. *Methods Ecol. Evol.* 9, 2155–2166. <https://doi.org/10.1111/2041-210X.13041>

Schulte to Bühne, H., Pettorelli, N., 2018. Better together: Integrating and fusing multispectral and radar satellite imagery to inform biodiversity monitoring, ecological research and conservation science. *Methods Ecol. Evol.* 9, 849–865. <https://doi.org/10.1111/2041-210X.12942>

Schweiger, A.K., Cavender-Bares, J., Townsend, P.A., Hobbie, S.E., Madritch, M.D., Kothari, S., Grossman, J.J., Gholizadeh, H., Wang, R., Gamon, J.A., 2020. Spectral niches reveal

743 taxonomic identity and complementarity in plant communities (preprint). BioRxiv.
 744 <https://doi.org/10.1101/2020.04.24.060483>

745 Smith, W.B., 2002. Forest inventory and analysis: a national inventory and monitoring program.
 746 Environ. Pollut. 116, S233–S242. [https://doi.org/10.1016/S0269-7491\(01\)00255-X](https://doi.org/10.1016/S0269-7491(01)00255-X)

747 Soberón, J., 2007. Grinnellian and Eltonian niches and geographic distributions of species. Ecol.
 748 Lett. 10, 1115–1123. <https://doi.org/10.1111/j.1461-0248.2007.01107.x>

749 Soria-Auza, R.W., Kessler, M., Bach, K., Barajas-Barbosa, P.M., Lehnert, M., Herzog, S.K.,
 750 Böhner, J., 2010. Impact of the quality of climate models for modelling species
 751 occurrences in countries with poor climatic documentation: a case study from Bolivia.
 752 Ecol. Model. 221, 1221–1229. <https://doi.org/10.1016/j.ecolmodel.2010.01.004>

753 Stavros, E.N., Schimel, D., Pavlick, R., Serbin, S., Swann, A., Duncanson, L., Fisher, J.B.,
 754 Fassnacht, F., Ustin, S., Dubayah, R., Schweiger, A., Wennberg, P., 2017. ISS
 755 observations offer insights into plant function. Nat. Ecol. Evol. 1, 0194.
 756 <https://doi.org/10.1038/s41559-017-0194>

757 Tilman, D., Clark, M., Williams, D.R., Kimmel, K., Polasky, S., Packer, C., 2017. Future threats
 758 to biodiversity and pathways to their prevention. Nature 546, 73–81.
 759 <https://doi.org/10.1038/nature22900>

760 Tsirogiannis, C., Sandel, B., 2016. PhyloMeasures: a package for computing phylogenetic
 761 biodiversity measures and their statistical moments. Ecography 39, 709–714.
 762 <https://doi.org/10.1111/ecog.01814>

763 Turner, W., 2014. Sensing biodiversity. Science 346, 301–302.
 764 <https://doi.org/10.1126/science.1256014>

- 765 Václavík, T., Meentemeyer, R.K., 2009. Invasive species distribution modeling (iSDM): Are
766 absence data and dispersal constraints needed to predict actual distributions? *Ecol.*
767 *Model.* 220, 3248–3258. <https://doi.org/10.1016/j.ecolmodel.2009.08.013>
- 768 Vilela, B., Villalobos, F., 2015. letsR: a new R package for data handling and analysis in
769 macroecology. *Methods Ecol. Evol.* 6, 1229–1234. [https://doi.org/10.1111/2041-](https://doi.org/10.1111/2041-210X.12401)
770 210X.12401
- 771 Watson, J.E.M., Keith, D.A., Strassburg, B.B.N., Venter, O., Williams, B., Nicholson, E., 2020.
772 Set a global target for ecosystems. *Nature* 578, 360–362. [https://doi.org/10.1038/d41586-](https://doi.org/10.1038/d41586-020-00446-1)
773 020-00446-1
- 774 Webb, C.O., Ackerly, D.D., McPeck, M.A., Donoghue, M.J., 2002. Phylogenies and Community
775 Ecology. *Annu. Rev. Ecol. Syst.* 33, 475–505.
776 <https://doi.org/10.1146/annurev.ecolsys.33.010802.150448>
- 777 Xiao, Z., Liang, S., Wang, J., Chen, P., Yin, X., Zhang, L., Song, J., 2014. Use of General
778 Regression Neural Networks for Generating the GLASS Leaf Area Index Product From
779 Time-Series MODIS Surface Reflectance. *IEEE Trans. Geosci. Remote Sens.* 52, 209–
780 223. <https://doi.org/10.1109/TGRS.2013.2237780>
- 781 Zurell, D., Zimmermann, N.E., Gross, H., Baltensweiler, A., Sattler, T., Wüest, R.O., 2020.
782 Testing species assemblage predictions from stacked and joint species distribution
783 models. *J. Biogeogr.* 47, 101–113. <https://doi.org/10.1111/jbi.13608>

FIGURE LEGENDS

Fig. 1. Implementation of the sequential procedure to construct and evaluate biodiversity model predictions. The first step defines the accessible area for each oak species. In this study, the accessible area is defined as the overlap of occurrences (confirmed presences) of sister species (a). The second step involves implementing an ensemble modeling approach for individual oak species but constrained to the accessible area established in the step one (b). Lastly, individual model predictions are stacked (S-SDMs) to obtain species richness, phylogenetic structure and composition predictions. These predictions are then evaluated using an independent dataset of species diversity and composition obtained from FIA (c).

Fig. 2. Bayesian Pearson's correlation coefficients between the predicted and observed species richness at different grain size. Observed grain size or FIA plots of $\sim 0.006 \text{ km}^2$ (a), 30 arcseconds or 1 km^2 aggregation (b), 0.25 degrees or $\sim 25 \text{ km}^2$ aggregation(c), and 0.5 degrees or $\sim 55 \text{ km}^2$ aggregation (d). Posterior distribution show ρ values from 4 sampling chains and 10,000 generations. The vertical black line represent the median (and 95% confidence interval, vertical gray dotted lines) ρ value from posterior distribution.

Fig. 3. Accuracy of assemblage composition predictions at different grain size. (a) Richness deviation, measured as the absolute deviation in assemblage richness predictions divided by the maximum observed FIA plot species richness, (b) prediction success or accuracy in assemblage

predictions, (c) True skill statistics, TSS, and (d) Sørensen similarity. Significant differences between stacking procedures (bS-SDM versus cS-SDM) were evaluated using Bayesian ANOVAs with FIA plots as random effect (significance levels for the Maximum A Posteriori p-values are: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, • $p < 0.1$). We also show Cohen's d effect size for each comparison. Note that the width of each boxplot is weighted by the sample size. Boxplots are ordered according grain size, from the observed grain size or FIA plot of ~ 0.006 km² (left-hand boxplots) to the largest grain size (right-hand boxplots), i.e., FIA plots aggregated at 0.5 degrees or ~ 55 km².

Fig. 4. Phylogenetic structure (standardized effect size–mean phylogenetic distance [SES-MPD]) of natural and predicted oak assemblages at four grain sizes. SES-MPD at observed grain size (~ 0.006 km²) (a), SES-MPD at 30 arcseconds or 1 km² aggregation (b), SES-MPD at 0.25 degrees or ~ 25 km² aggregation(c), and SES-MPD at 0.5 degrees or ~ 55 km² aggregation (d).

Fig. 5. Bayesian Pearson's correlation coefficients between the predicted (bS-SDM and cS-SDM) and observed assemblage phylogenetic structure at different grain size. (a-c) Observed grain size or FIA plots of ~ 0.006 km², (d-f) 30 arcseconds or 1 km² aggregation, (g-i) 0.25 degrees or ~ 25 km² aggregation, and (j-l) 0.5 degrees or ~ 55 km² aggregation. Posterior distributions show ρ values from 4 sampling chains and 10000 generations. The vertical black line represent the median (and 95% confidence interval, vertical gray dotted lines) ρ value from posterior distribution.

Fig. 1.

834 **Fig. 2.**

835

836

837

Fig. 3.

842 **Fig. 4.**

843

844

845 **Fig. 5.**

846

847

Supplementary material

Fig. S1. Model accuracy comparison for each algorithm and the resulting ensemble model for the 49 oak species. AUC = area under the ROC curve, TSS = true skill statistics, BRT = Boosted Regression trees, GAM = Generalized Additive Model, GLM = Generalized Linear Model, RF = Random Forest.

855

856 **Fig. S2.** Distribution of oak tree prevalence levels for the four grain sizes estimated from the FIA
857 dataset.

858

859