Forest Ecology and Management 539 (2023) 120972

Contents lists available at ScienceDirect

Forest Ecology and Management

journal homepage: www.elsevier.com/locate/foreco

Improved prediction of tree species richness and interpretability of environmental drivers using a machine learning approach

Lian Brugere [a], Youngsang Kwon [a,*], Amy E. Frazier [b], Peter Kedron [b]

[a] Department of Earth Sciences, University of Memphis, Memphis, TN 38152, USA
[b] School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, AZ 85281, USA

ARTICLE INFO

Keywords:
Tree species richness modeling
Generalized linear model
Random forest
Neural networks
Deep learning
FIA

ABSTRACT

Biodiversity is in decline globally and predicting species diversity is critically important if current trends are to be reversed. Tree species richness (TSR) has long been a key measure of biodiversity, but considerable uncertainties exist in current models, particularly given the classic statistical assumptions and poor ecological interpretability of machine learning outcomes. Here, we test several ecologically interpretable machine learning approaches to predict TSR and interpret the driving environmental factors in the continental United States. We develop two artificial neural networks (ANN) and one random forest (RF) model to predict TSR using Forest Inventory and Analysis data and 20 environmental covariates and compare them to a classic generalized linear model (GLM). Models were evaluated on an independent, unseen testing dataset using R^2 and Mean Absolute Error (MAE) and residual spatial autocorrelation analysis. An Interpretable Machine Learning approach, SHapley Additive exPlanations (SHAP), was adopted to explain the major environmental factors driving TSR. Compared to a baseline GLM ($R^2 = 0.7$; MAE = 4.7), the ANN and RF models achieved R^2 greater than 0.9 and MAE<3.1. Additionally, the ANN and RF models produced less spatially clustered TSR residuals than the GLM. SHAP analysis suggested that TSR is best predicted by Aridity Index, Forest Area, Altitude, Mean Precipitation of the Driest Quarter and Mean Annual Temperature. SHAP further revealed a non-linear relationship of environmental covariates with TSR and complex interactions that were not revealed by the GLM. The study highlights the need for conservation efforts of forest areas and reducing precipitation-related physiological stress on tree species in low forested but arid regions. The machine learning approach used here is transferrable for studies of biodiversity for other organisms or prediction of TSR under future climatic scenarios.

1. Introduction

Biodiversity is a critical ecological indicator currently undergoing significant global declines, which are adversely impacting humanity (Cardinale et al., 2012). For decades, the cornerstone of biodiversity science has been understanding the drivers and consequences of variation in species richness and how abundances change across spatial and temporal scales (MacArthur et al., 1972; Rosenzweig, 1995). Despite a shift in focus over the last three decades to approaches that prioritize functional diversity (Swenson et al., 2011), species-based approaches remain relevant and have given rise to some of the most influential ecological theories (MacArthur and Wilson, 1967; Tilman, 1982; Hubbell, 2001). Predicting biodiversity trends is critically important to understand the impact of global climate change (Pereira et al., 2013) and regional land use and land cover change (Cardinale et al., 2012) on ecosystems and humans.

Tree species richness (hereafter TSR) is the number of distinct tree species represented in an ecological community, landscape, or region. TSR has long been an important component of biodiversity studies (Gentry, 1988; Cleland, 2011), particularly with respect to global change (Iverson and Prasad, 2001). Recent studies have shown that richness increases with higher stand-level productivity in forests (Huang et al., 2018), interactions of higher trophic levels (Schuldt et al., 2017), forest stability (Ouyang et al., 2020), and nutrient cycling and soil-related processes (Haghverdi and Kooch, 2019). Therefore, accurately predicting and modeling broad-scale changes in TSR is critical to forecasting where declines may occur in the future and what effects those changes may have on other species or ecosystem services.

However, modeling and predicting TSR at broad spatial scales (e.g., continental to global) has been limited by both data availability and

* Corresponding author.
E-mail address: ykwon@memphis.edu (Y. Kwon).

https://doi.org/10.1016/j.foreco.2023.120972
Received 15 May 2022; Received in revised form 8 February 2023; Accepted 26 March 2023
Available online 21 April 2023

methodological concerns. The lack of systematic species occurrence data with consistent spatial grain at continental to global scales has been well-documented (Belmaker and Jetz, 2010; Keil and Chase, 2019). In North America, the USDA Forest Inventory and Analysis (FIA) data has been particularly effective for modeling TSR. The FIA data provide a consistent, systematic, annual sample of plot-level tree information in the United States (Bechtold and Patterson, 2005). The data have been used in biodiversity studies (Woodall et al., 2010; Zhu et al., 2015), but most previous work has focused on regional or subcontinental-level analyses of TSR (e.g., across the eastern United States; Fan and Waring, 2009; Kwon et al., 2018). The nation-wide fixed-radius plot design FIA inventory offers the opportunity for larger scale estimates of TSR across the entire United States, but the large dataset (millions of observations) has become unwieldy for classic statistical modeling approaches.

Methodological concerns have also limited the development of broad-scale predictions of TSR. Most prior studies have used correlative (or linear) models to predict TSR, with the most common approach being a general linear model (Sarr et al., 2005). However, general linear models have been found to be inappropriate for TSR prediction because the spatial patterns of TSR are driven by complex, non-linear environmental relationships, and TSR is not typically normally distributed (Wang et al., 2011; Kwon et al., 2018). To address the non-normality, generalized linear modeling (GLM) has been used (Wang et al., 2011; Kwon et al., 2018), but the collinearity of predictor variables and spatial autocorrelation of residuals present challenges. Variable selection methods, such as Least Absolute Shrinkage and Selection Operator (LASSO), have been used to reduce multicollinearity (Kwon et al., 2018), and spatial autoregressive models (Svenning and Skov 2007) or eigenvector spatial filtering models (Kwon et al., 2018) have been employed to account for spatial autocorrelation. However, issues remain with the ability of these models and techniques to capture the complex spatial patterns of TSR with highly interactive environmental covariates.

Data-driven, machine learning models have shown high prediction accuracy when applied to global ecology and biodiversity studies (Christin et al., 2019; Maina, 2021). Compared to classic predictive models, machine learning models provide several benefits. First, machine learning models use general-purpose learning algorithms to find patterns in big and unwieldy data (Bzdok et al., 2017) while classic predictive models tend to overfit when many covariates are included. Second, machine learning methods make minimal assumptions about data structure or data-generating systems, and they can be effective even when the underlying systems are unknown, difficult to describe, have complex interactions or the sample data are noisy (Liu et al., 2018). This loose assumption about data structure is relevant for TSR studies because the forces driving species richness spatial patterns are multi-dimensional, non-linear, and highly correlated (Li et al., 2017). Third, in contrast to classic statistical approaches where the entire data sample is considered during hypothesis testing (Wasserstein and Lazar, 2016), with machine learning approaches, the sample data are typically separated into subsets for training, validation, and testing. The algorithm is fit with the training data, and hyperparameters are tuned on the validation data to produce the best model with the highest prediction accuracy. The fitted model is then empirically evaluated using the testing set. With machine learning, the desired relevance of a statistical relationship in the underlying population is ascertained by explicit evaluations on new data rather than formal mathematical proofs as it is for classic regression (Breiman, 2001; Wasserstein and Lazar, 2016). This feature ensures the machine learning model not only produces low bias but also low variance (i.e., high generalization power) on new or unseen data from the modeled system.

Among machine learning models, random forests (RF, Breiman, 2001) and artificial neural networks (ANNs, Goodfellow et al., 2016) have been widely applied for biodiversity assessments as they are robust nonparametric, non-linear learners that have demonstrated good predictive power (Bland et al., 2015; Wu and Liang, 2018). RF and ANNs models are also fairly insensitive to problems of multicollinearity, which makes them well suited for species prediction models when a large number of environmental covariates are being tested, eliminating the need for a variable pre-selection step such as LASSO. In addition, both RF and ANNs are stable and can tolerate outliers or noisy data (Liu et al., 2018). While interpretability of ML outputs is a major concern in ecological studies (Welchowski et al., 2021), Interpretable Machine Learning (IML) has been offered to explain increasingly complex machine learning models to provide insights on the modeled system and predictors (Molnar 2020; Welchowski et al., 2021). A recent review by Linardatos et al. (2021) suggested the Shapley Additive exPlanations (SHAP, Lundberg and Lee, 2017) method as the most comprehensive, model-agnostic method for its versatility. These developments offer an opportunity to examine environmental relationships in greater depth.

This study develops a machine learning approach for predicting TSR at broad spatial scales to improve prediction accuracy and determine the environmental factors driving TSR. Specifically, we test two approaches – RF and ANNs – for predicting TSR in the continental United States with the nation-wide FIA database and 20 widely-studied environmental covariates. We apply the SHAP IML method to analyze the importance of these covariates and unveil their relationships to TSR. We compare these models to the classic GLM approach. The work is novel in utilizing ML models to improve the geographic prediction accuracy of TSR and adopting IML method to explain the importance of environmental covariates and their relationships to TSR.

2. Data and variable construction

2.1. Forest Inventory and analysis database

The FIA program provides consistent, nationwide tree census information on the extent, condition, status, and trends of United States forest resources (Bechtold and Patterson, 2005). Data are collected via a systematic, five-year rolling annual inventory system with a unified, fixed-radius plot design. Within each plot, adult trees (>=5.0 in. or 12.7 cm diameter at breast height, DBH) are tallied in four, 24-foot or 7.3 m fixed-radius subplots, and saplings (between 1.0 in. or 2.5 cm and 5.0 in. or 12.7 cm DBH) are tallied in four, 6.8-foot or 2.1 m fixed-radius microplots for core attributes such as tree diameter, height, damage, and forest type (Bechtold and Patterson, 2005). Plot locations on private land are generally swapped for privacy protection (McRoberts et al., 2005), but the up to 0.8 km shift in location is negligible in studies with large spatial extent (Woodall et al., 2010). We retrieved the FIA database (version 1.8.0.00) for the continental United States (a.k.a. the lower 48 states) from the FIA DataMart (https://apps.fs.usda.gov/fia/datamart/) for the period 2013 to 2018. We include both subplots and microplots, with all data from each plot aggregated into a grid (described in section 2.2). The dataset comprises 143,810 plots, which include 1,151,062 observations consisting of 714,805 adult trees and 436,257 seedlings.

2.2. TSR outcome variable

To calculate a standard TSR, we overlaid a 20 × 20 km grid (a total of 20,251 cells) over the continental United States and mapped each FIA plot to this grid (cf. Kwon et al., 2018). Among these grid cells, 15,310 contained at least one FIA subplot or micro-plot; the remaining 4,941 grid cells did not contain any plots and were eliminated from the analysis. TSR was calculated as the number of distinct tree species for each grid unit (Fig. 1a). The 20 × 20 km analytic unit is chosen to capture spatial variation of TSR and environmental factors across continental US. A previous study (Kwon et al. 2018) found that TSR saturation occurs at around 20 FIA plots, with the FIA sampling intensity being one plot per 6000 acres (24.3 km2) of forest, there are 17–18 FIA plots per 20 km × 20 km of forested area.

The total number of tree species from FIA samples in the continental United States is 390. Across the grid, TSR ranges between 1 and 60, with

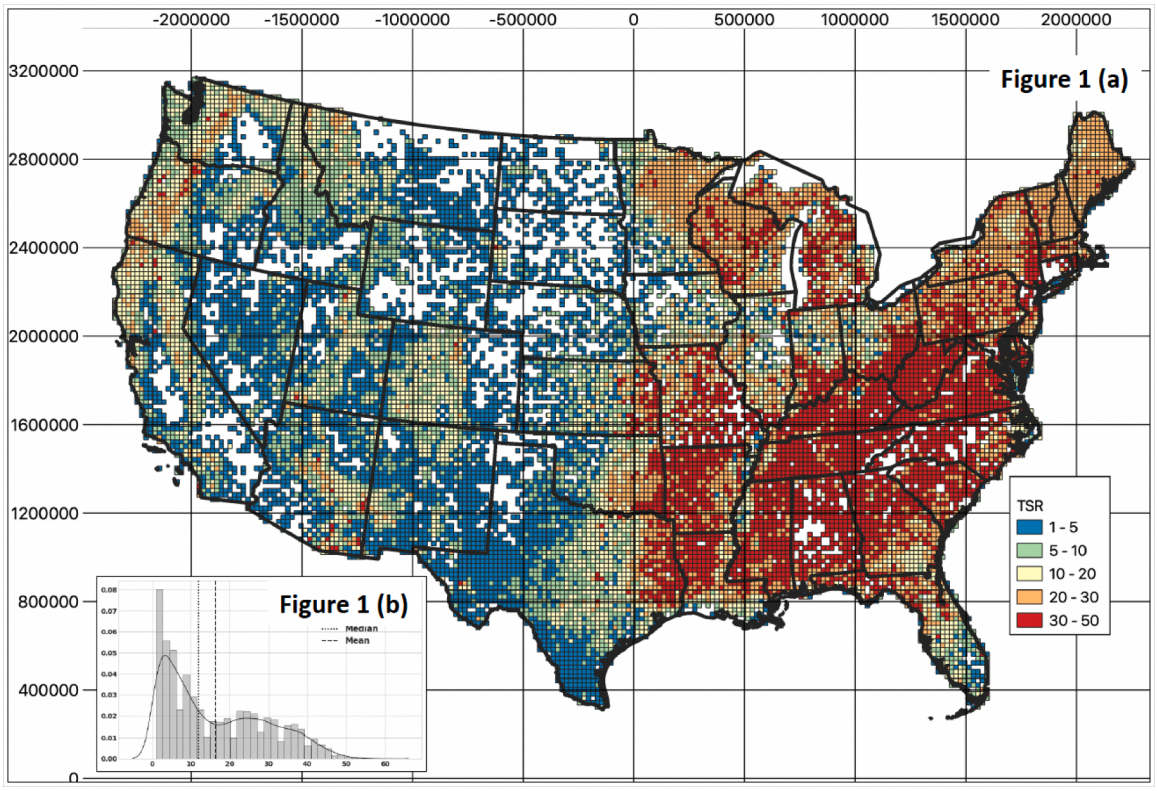

Fig. 1. (a) Tree species richness (TSR) calculated from USDA Forest Inventory and Analysis (FIA) database aggregated to a 20 km by 20 km grid system for the continental United States (lower 48 states), (b) TSR frequency distribution and kernel density estimate (KDE) point observations.

a median of 12. TSR varies considerably across the U.S., and the distribution is bimodal and positively skewed (Fig. 1b). In general, higher TSR values are in the central and southeast U.S. and lower values are located in the west.

2.3. Environmental covariates

Twenty environmental covariates identified in prior TSR studies (Fan and Waring, 2009; Wang et al., 2011; Kwon et al., 2018; Kwon et al., 2019) were chosen as predictor variables. To remain consistent with recent studies predicting TSR using FIA data, we grouped these variables into seven categories following Kwon et al. (2018) (Table S1). The soil hydrological group (SHG) variable is the only categorical variable and is converted to continuous via one-hot encoding. All predictor variables are standardized to [0,1]. For additional details on the data sources of the covariates, readers are directed to supporting document S1.

3. Methods

Upon the construction of environmental covariates and TSR variable,

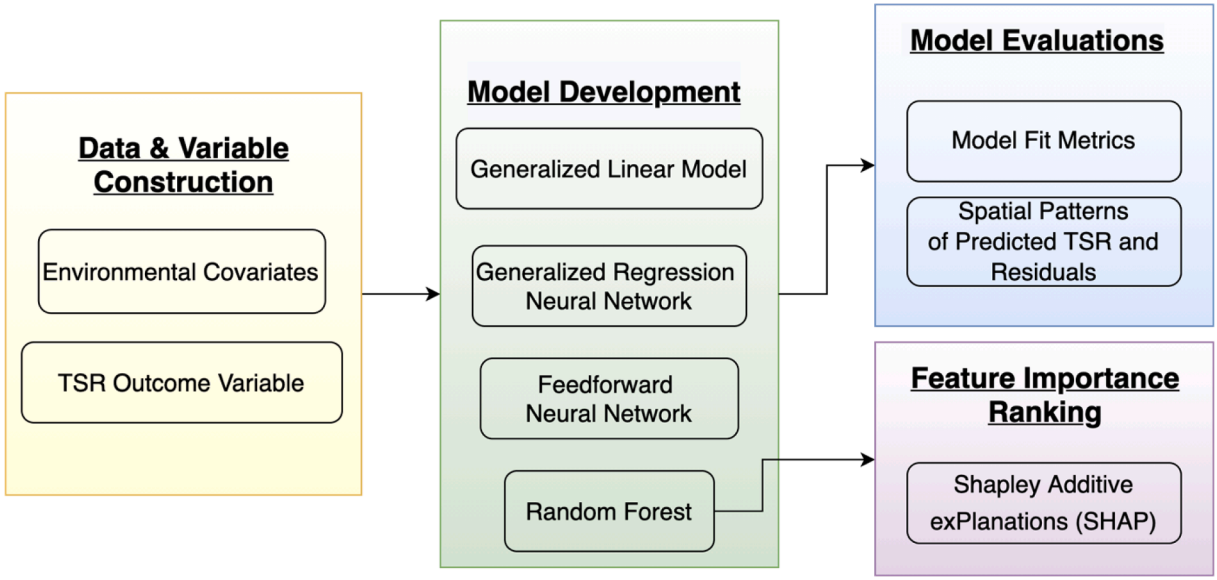

Fig. 2. Methodology flowchart.

we develop four types of models for TSR prediction: a generalized linear model (GLM), random forest (RF), and two artificial neural network (ANN) models – a generalized regression neural network (GRNN) and a feedforward neural network (FFNN) (Fig. 2). First, we develop a GLM model using the entire dataset (15,310 grid cells with 20 covariates) and examine the influence of each predictor variable on TSR. Second, we develop the RF, GRNN and FFNN models using training and validation sets. For the RF and two ANNs, the dataset is randomly split into three folds: 1) 80% of the cells are used to train the models, 2) 10% of the cells are used for validation to provide an unbiased evaluation of a model fit on the training dataset while tuning model architecture and hyperparameters if present, 3) 10% of the cells are used for testing to provide an unbiased evaluation of a final model fit after parameter tuning. To evaluate model performance, we compare fit metrics of the four models computed on the testing dataset and geographic patterns of predicted TSR in different regions across the United States. Lastly, we calculate a mean SHAP value for each environmental covariate in the RF model to rank variable importance for predicting TSR. Details of each model and the SHAP analysis are provided below. See Fig. 2 for the complete methodology flowchart and reproducible codes are available at https://github.com/lydiabrugere/tsrmodel.

3.1. Model development

The four models differ in terms of their characteristic features (Table 1). This section provides model specifications such as architectures and hyperparameters of trained models.

3.1.1. Generalized linear model (GLM)

The GLM is a baseline model with a logarithmic link function and negative binomial residuals. As species richness values are counts, appropriate statistical families for the error distribution are the poisson or negative binomial distribution. To choose an appropriate model family, we run GLM with a log link function and poisson-distributed residuals for TSR based on all 20 environmental covariates. Since the variance and mean of residual TSR were unequal, a negative binomial error distribution with a log-link function was chosen for the baseline GLM. To assess the explanatory power of each predictor variable, we report standardized regression coefficients of each predictor. GLM has a fixed architecture with no optimization method, thus no model training is done (Table 1). The model is developed using the python package *statsmodels* (Seabold and Perktold 2010).

3.1.2. Random forest (RF)

RF offers more flexibility than GLM but is more complex to develop (Table 1). Several hyperparameters in RF models can be trained for model optimization, such as the number of decision trees. Sensitivity tests have been carried out in previous studies to determine the optimal values for these hyperparameters in RF models. Liaw and Wiener (2001) recommend using the number of predictors divided by three as the value for the number of variables used at each split for regression. However, it is often feasible to programmatically train for optimal values in a modeled system. We trained the following hyperparameters with 10-fold cross-validation (final trained specifications are reported in parenthesis): 1) number of trees (200), 2) maximum depth of the tree (68), 3) minimum number of samples required to split an internal node (10), 4) minimum number of samples required to be at a leaf node (5), and 5) number of features that can be searched at each split (4). We implemented the RF model using the python package *sklearn* (Pedregosa et al., 2011).

3.1.3. Generalized regression neural network (GRNN)

GRNN is a single-pass ANN consisting of a fixed architecture of four layers: input, pattern, summation, and output layers (Specht, 1991). GRNN is a probabilistic neural network, and the additional knowledge needed to fit a GRNN model is relatively small (Specht, 1991). The input layer fully connects the raw input to the pattern layer, of which the function is a Radial Basis Function, typically the Gaussian kernel function. The width of Radial Basis Function, also known as the spread constant σ, is the only unknown parameter in GRNN models. The training of a GRNN model is essentially to determine the optimum value of σ. In general, if the input feature values are high, σ is high, and vice versa. GRNN is sensitive for cases when input feature values are of various ranges, which is the case with the variables being used to predict TSR here. Thus, we normalized the input data [0,1] and tuned the value for σ to [0.05, 1] with a step of 0.005, and evaluated the performance of σ based on cross-validated mean absolute error value on the validation dataset. The optimal value of the spread constant σ was 0.115 from the best performing GRNN model. GRNN is implemented using the python package *neupy* (Shevchuk, 2015).

3.1.4. Feedforward artificial neural network (FFNN)

FFNN is one of the most common ANNs (Lek et al., 1996). FFNN is typically best suited for tabular data and is a popular choice for regression tasks (Table 1). FFNN consists of at least three layers: an input layer accepting raw data, one or more hidden layers used for transformations, and an output layer for prediction. In FFNN models, performance is highly influenced by the architecture and hyperparameters (more options compared to GRNN), and training relies heavily on the data and modeled system itself. To obtain the best performing FFNN model, we trained the following hyperparameters and selected optimal values based on TSR prediction performance on the validation dataset: (1) number of hidden layers (1–5), (2) number of neurons in each layer (256, 512 or 1024), (3) dropout rate (0–1), i.e., a regularization technique to avoid ANNs overfitting (Srivastava, et al., 2014) (4) optimizer ('adam', 'sgd' or 'rmsprop'), and (5) batch size (128, 256 or 512). A ReLU transfer function (Agarap, 2018) is used for all hidden and output layers to enable the network to learn nonlinear relationships with computational efficiency. The versatile and computationally efficient 'adam' optimizer (Kingma and Ba, 2015) was chosen to train the

Table 1
Comparison of model features (GLM: Generalized Linear Model; RF: Random Forest; GRNN: Generalized Regression Neural Network; FFNN: Feedforward Neural Network).

Features	GLM	RF	GRNN	FFNN
Data Types	Tabular	Tabular	Tabular, Image, Audio	Tabular, Image, Audio
Number of Data Points	Low	Moderate	Low	High
Constraints on Data Structure	High	Low	Moderate	Low
Prediction Task	Regression	Regression or Classification	Regression	Regression or Classification
Implementation Complexity	Low	Moderate	Low	High
Architecture to Tune	No	Yes	No	Yes
Number of Tuning Parameters/Hyperparameters	Low	Moderate	Low	High
Interpretability	High	Moderate	Low	Low
Training Time	Low	Moderate	Low	High
Optimization Method	No	Yes	No	Yes
Predictive Accuracy	Low	High	High	High

network. The model was trained for 1,000 epochs on the training dataset, and the training batch size was 128. The final, trained FFNN model consisted of three hidden layers: the input layer and the third hidden layer had 1,024 neurons, and the first and second hidden layers had 512 neurons. Dropout rates for the three hidden layers were 0.46, 0.50 and 0.59 respectively. FFNN was implemented using the python packages *Keras* (Chollet et al., 2015) and *Tensorflow* (Abadi et al., 2015), and hyperparameter tuning was conducted using *hyperopt* (Bergstra et al., 2015).

3.2. Model evaluation

Prediction accuracy for all models is assessed through model fit metrics including the coefficient of determination (R^2) and mean absolute error (MAE). We also assess residual spatial autocorrelation to understand its impact on the model reliability, and this is accomplished in two steps. First, for each grid, the standardized Pearson residual is calculated as its raw residual divided by the standard error of all residuals. Second, residuals for all 15,310 grids are evaluated using Moran's I. In addition, TSR prediction accuracy is assessed using a non-parametric Wilcoxon signed-rank test, in which TSR predictions made by all four models are compared with FIA-based TSR observations paired at the grid level. This test is done as TSR data are not normally distributed. The null hypothesis is that the median difference of TSR between predictions and FIA observations of matched grids is zero.

3.3. Feature importance

Shapley Additive exPlanations (SHAP, Lundberg and Lee, 2017) is a state-of-the-art IML method based on the concept of Shapley values from cooperative game theory (Shapley, 1951). SHAP has been benchmark tested in machine learning models as a more unified feature attribution method in terms of model consistency and accuracy compared to other widely used approaches such as permutation feature importance and Gini importance (a.k.a. embedded feature importance in random forest) (Lundberg and Lee, 2017). SHAP values are calculated as the average difference between the predicted value and a baseline value. The absolute value of SHAP is a measure of how much impact a model has on predicting TSR; the higher the SHAP value, the more impact. The sign of the SHAP value indicates which direction a predictor drives TSR.

SHAP has been widely applied to different fields since its initial implementation (Linardatos et al., 2021). As a model agnostic approach, SHAP can practically explain any kind of machine learning model, but the computational complexity grows exponentially with model complexity. It is less computationally expensive to explain an RF model than a deep neural network model. A general rule when implementing SHAP is to explain a pre-trained model with good performance, i.e., low bias and low variance. Lundberg and Lee (2017) demonstrated the accuracy and consistency of SHAP implementation of tree-based models. For these reasons, we apply SHAP to explain how our trained RF model uses environmental covariates to make TSR predictions and compare the relative contribution of each environmental covariate to TSR. We also report the standardized regression coefficients from the GLM model for comparison of feature importance.

4. Results

4.1. Modeling results

Of the four models tested, the GLM model has the lowest testing R^2 (0.739) and highest MAE (4.713). The GLM also has residual Moran's I value (0.445) three to four times higher than all other models (Table 2). Correspondingly, the predicted median TSR value from the GLM for the testing set (13.69) was higher than the true median (12) and higher than the predictions from any other model (Table 2). The GLM does not require model training and therefore is not included in training-based comparisons.

The three remaining models all achieved testing R^2 higher than the GLM. Each of these models also permit the calculation of R^2 for the training set. The RF model showed the highest R^2 (0.955) and lowest MAE (1.931) on the training dataset and the second highest R^2 (0.903) and MAE (2.904) on the test dataset (Table 2). The RF model also achieved the lowest Moran's I residual (0.083) (Table 2), and its predicted median TSR value (12.59) was the second closest to the true median (12). The GRNN model achieved the second highest R^2 (0.932) and second lowest MAE (2.195) in the training dataset but the lowest R^2 (0.886) and highest MAE (3.142) after GLM in the test dataset (Table 2). The GRNN model also attained the second lowest Moran's I residual (0.134), and the predicted median TSR value (12.72) was higher than the true value (12) (Table 2). The FFNN model showed the lowest R^2 (0.917) and highest MAE (2.505) in the training dataset, whereas it achieved the highest R^2 (0.907) and lowest MAE (2.851) in the test dataset. The FFNN model obtained the second lowest residual Moran's I (0.124). The predicted median value of TSR was 12.29 in the FFNN model, and this was the only model that achieved a not significantly different median TSR value compared to the observed median (Table 2). Collectively, these results suggest that the RF model outperformed the GLM and the two neural networks for predicting TSR.

4.2. Feature importance

We examined which features were most important for predicting TSR across the United States. First, as a baseline for comparisons with the widespread use of GLMs in prior studies, we present feature importance from our fitted GLM. From the GLM, we used the estimated standardized regression coefficients to identify forest area, mean annual precipitation, range of mean annual temperature, annual range of temperature, and altitude as the five most important predictors of TSR. Except for altitude, all of these predictors positively influenced TSR (Table 3).

Second, we identified the importance of features in our best fitting RF model using the SHAP method. The SHAP summary plot (Fig. 3) displays the relative importance of the 20 environmental covariates included in the RF model in descending order by their absolute mean SHAP values over the entire gridded dataset. This analysis reveals that the five most important variables for predicting TSR are (in descending order): aridity index, forest area, altitude, mean precipitation of the driest quarter, and mean annual temperature. The vertical dispersion for each predictor in Fig. 3 represents a mixed effect with other predictor variables on TSR. If a predictor had no interaction with another

Table 2
Model performance comparisons between the Generalized Linear Model (GLM), random forest (RF), generalized regression neural network (GRNN) and feedforward neural network (FFNN).

Model	Train		Testing		Residual Moran's I	Z-score of Moran's I	Continental United States (Median = 12.0, N = 15310)		
	MAE	R^2	MAE	R^2			Median	W*	P-value (two-tailed)
GLM	–	–	4.713	0.739	0.445	121.482***	13.69	55,122,841	0.006
RF	1.931	0.955	2.904	0.903	0.083	18.651***	12.59	57,574,340	0.060
GRNN	2.195	0.932	3.142	0.886	0.134	29.988***	12.72	57,483,148	0.040
FFNN	2.505	0.917	2.851	0.907	0.124	27.779***	12.29	57,211,620	0.903

Significance level: ***: $p < 0.001$; W* is the sum of the ranks of the differences (positive)

Table 3
GLM-based relations between the 20 predictor variables and tree species richness (bold indicates covariates with the highest standardized coefficients).

Category	Abbreviation	Variable	Standardized Coefficient	Standard Error
Intercept			2.448	0.003**
Areal factors	**FA**	**Forest Area (km^2)**	**0.394**	0.013***
	WA	Water Area (km^2)	0.020	0.012
Climatic seasonality	**ART**	**Annual Range of Temperature**	**0.285**	0.083***
	PSN	Precipitation Seasonality	−0.016	0.021
	TSN	Temperature Seasonality	0.008	0.098
Energy availability	MAT	Mean Annual Temperature (°C)	0.045	0.075
	MTWQ	Mean Temperature of Warmest Quarter (°C)	−0.023	0.062
	PET	Potential Evapotranspiration (mm)	−0.117	−1.738
Energy-water dynamic	EWD	PET-PET^2 + MAP	0.239	0.051***
Habitat heterogeneity	RA	Range of Altitude	−0.204	0.042***
	RMAP	Range of Mean Annual Precipitation	−0.016	0.016
	RMAT	**Range of Mean Annual Temperature**	**0.337**	0.036***
Limiting climatic factors	MFDF	Mean Frost Day Frequency	−0.095	0.035**
	MPDQ	Mean Precipitation of Driest Quarter	0.236	0.034***
	MTCQ	Mean Temperature of Coldest Quarter	0.028	0.089
	ALT	**Altitude**	**−0.275**	0.033***
	SHG	Soil Hydrological Group	−0.013	0.009
Water availability	AET	Annual Evapotranspiration (mm)	−0.003	0.011
	AI	Aridity Index	−0.238	0.034***
	MAP	**Mean Annual Precipitation**	**0.349**	0.042***
Null Deviance		15,289	Residual Deviance (% explained)	3142 (79.45)

Significance level: ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$

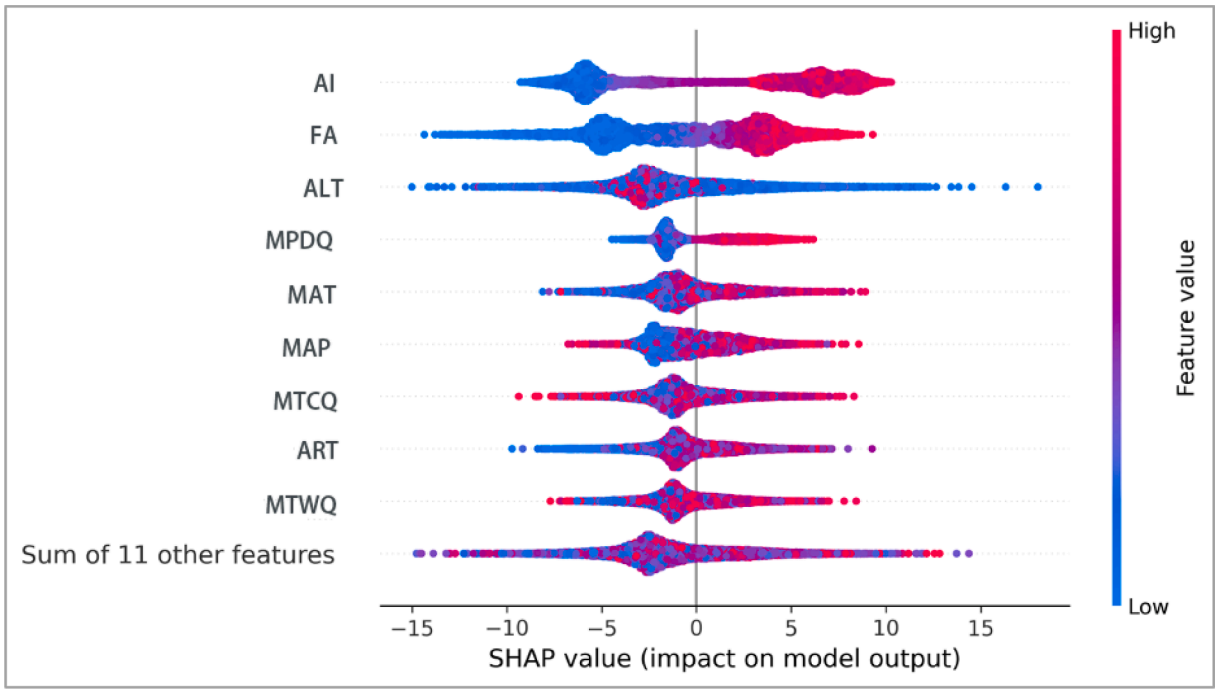

Fig. 3. SHapley Additive exPlanations (SHAP) summary plot from the best fitting RF model of the relative importance of 20 environmental covariates in the prediction of tree species richness (TSR) (The horizontal location on x-axis indicates whether an observation of a predictor has a negative/positive and high/low impact on the prediction of TSR; Vertical dispersion for each variable represents its interaction effects with other features). Refer to Table 1 in S1 for the full name of each predictor variable.

predictor, the plot for that predictor would be flat. Some level of interaction with other variables was observed for all covariates in Fig. 3. Overall, a higher aridity index (AI, i.e., greater humidity) and larger forest area (FA) were found to drive higher TSR, while a drier climate and lower forest area was found to drive lower TSR. Similarly, mean precipitation of the driest quarter (MPDQ) and mean annual temperature (MAT) were also found to have a positive relationship with TSR. For altitude (ALT), the impact on TSR was mixed, but higher altitude values (red dots in Fig. 3) in some grids had a negative impact on TSR.

We next used the SHAP measure to examine how the identified important predictors impact TSR prediction and their interactions with other predictors (Fig. 4). The absolute SHAP value means the magnitude of impact a predictor has on TSR, and a positive SHAP value means a positive impact on TSR while a negative value means a negative impact. We found that the three covariates representing precipitation availability all have a non-linear relationship with TSR (Fig. 4a, 4d, 4f). In arid regions, which are widely represented across the western United States (Figures S2.1, S2.2, S2.3, Table S2.1), precipitation availability had a negative impact on the predicted TSR. As aridity (AI) and mean annual precipitation (MAP) increase, the magnitudes of their impacts decrease in a near-linear fashion. Both AI and MAP have the strongest negative impact on TSR when they are lowest.

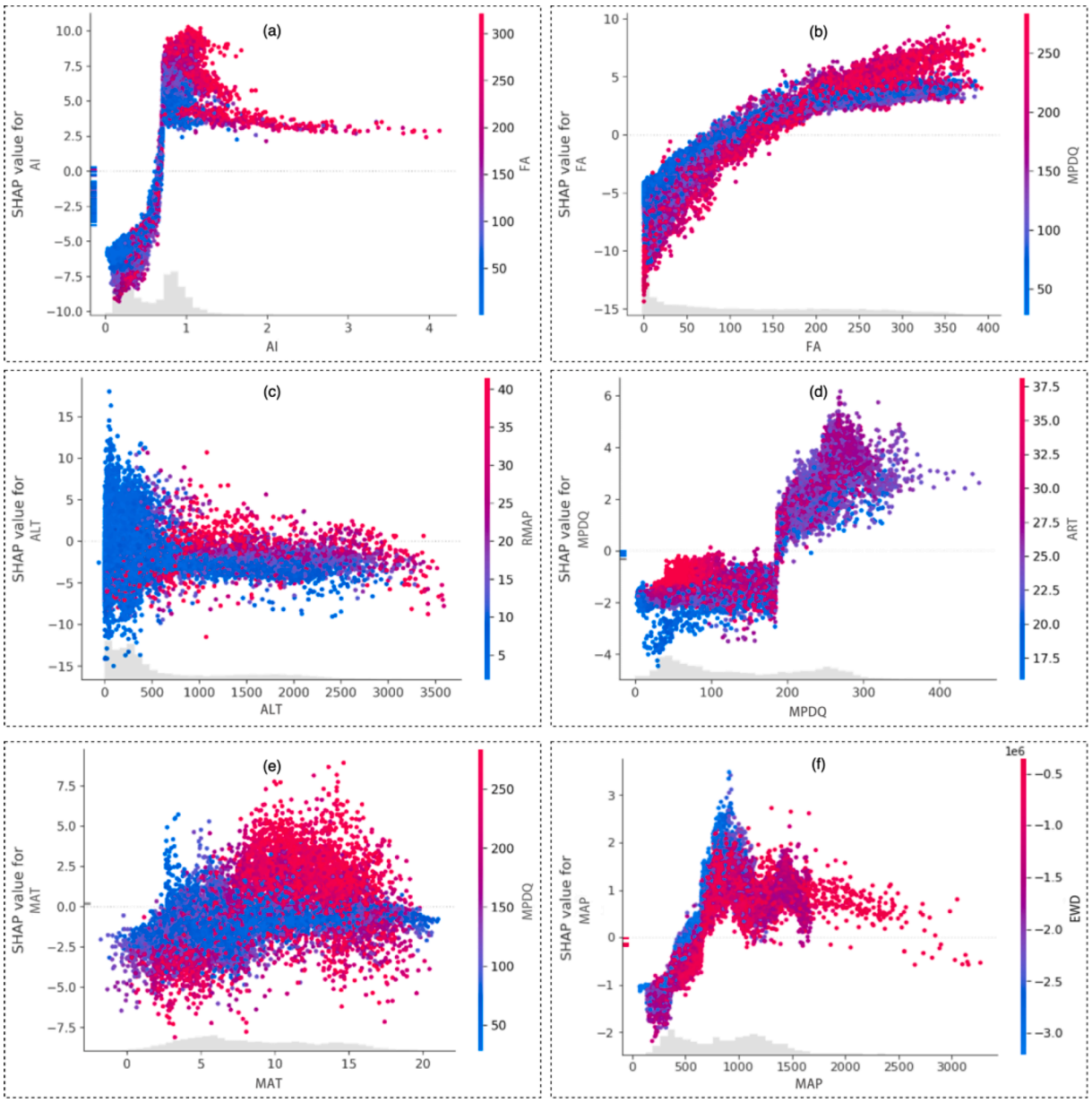

Fig. 4. SHapley Additive exPlanations (SHAP) values for the six most important environmental covariates in the prediction of tree species richness (TSR) (the color is denoted by each predictor's most interactive predictor -high (red) or low (blue); the vertical dispersion in each plot shows that the same value for the predictor can have a different impact on the TSR for different grids. This means there are non-linear interaction effects in the model between the environmental covariate and other predictors. Histogram of each covariate is shown in the grey plots along the x-axis.

In more humid regimes, such as the eastern U.S., precipitation availability contributes positively to predicted TSR. As the AI, MPDQ, and MAP continue to increase (i.e., as the climate becomes more humid), the magnitude of positive impacts on TSR becomes greater until the effect is compounded by interactions with other covariates (Fig. 4a, 4d, 4f). For instance, when the aridity index is greater than 0.9 (e.g., in the southeast U.S.), it interacts strongly with forest area. The greater the forest area, the stronger the impact of the aridity index on TSR until the impact becomes stable as forest area approaches the maximum cell value of 400 km^2 (Fig. 4a). In humid areas, MPDQ interacts with the annual range of temperature (ART) (Fig. 4d), and MAP interacts with Energy-Water Dynamic (expressed as $PET\text{-}PET^2 + MAP$) (Fig. 4f). Both interactions are non-linear and complex with no patterns shown. This mixed effect of environmental energy and precipitation on TSR agrees with previously proposed relationships between TSR and climate for angiosperms (Francis and Currie, 2003).

Forest area (FA) was the second strongest predictor of TSR. The impact of forest area on TSR (Fig. 4b) resembles the species-area relationship identified in Preston (1962). The continental United States has very high forest area variability ranging from 0 to 395 km^2 per grid cell (400 km^2) with a standard deviation of 107 km^2 (S2 Fig. 4). Areas with low forest area were predicted to have low TSR in all four models; a similar relationship is observable in Little's range maps (e.g., Montoya et al., 2007). As forest area increases, the impact of area on TSR levels off, and forest area interacts strongly with MPDQ. Overall, the higher the MPDQ, the greater the mixed impact of forest area and MPDQ on TSR in highly forested areas.

Altitude was the third strongest predictor of TSR, and it impacts TSR in a non-linear manner and interacts strongly with the range of mean annual precipitation (RMAP) when altitude is relatively high (Fig. 4c). In lowlands (i.e., altitude $<$ 500 m, most of Eastern US), the impact of altitude on TSR varies considerably and ranges from a strong negative to

a strong positive. This finding indicates a compounding impact from other environmental covariables on TSR. As altitude increases, its average impact on TSR becomes marginal but stays largely negative, and the range of MAP (RMAP) starts to mediate the negative impact of altitude on TSR. The relationship between altitude and TSR found in this study is consistent with prior studies that vascular plant species richness does not maximize at a certain altitudinal zone in the Himalaya (Vetaas and Grytnes, 2002).

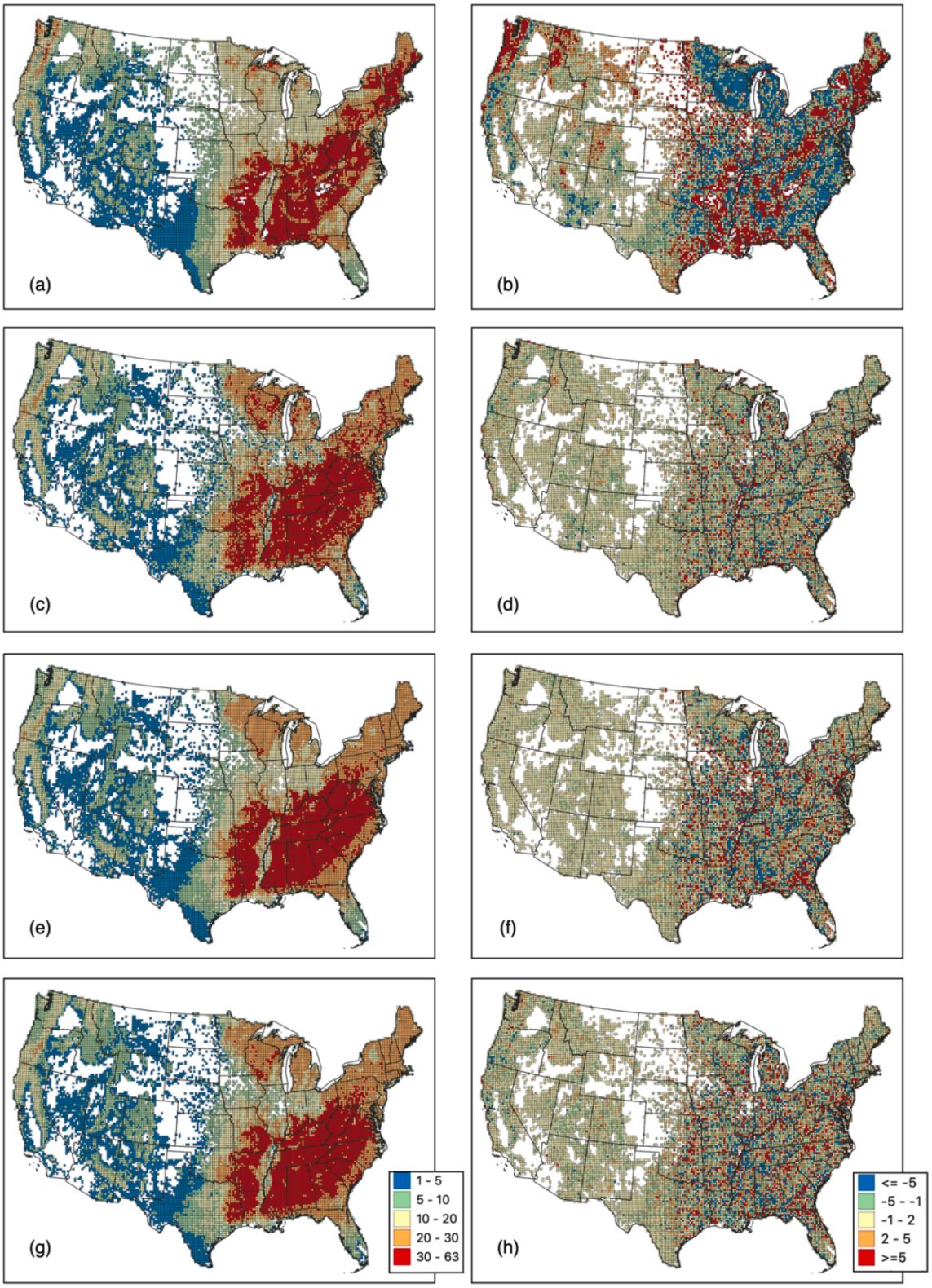

Fig. 5. Generalized Linear Model (GLM) prediction (a), GLM residuals (b), Random Forest prediction (c), Random Forest residuals (d), Generalized regression neural network (GRNN) prediction I, GRNN residuals (f), feedforward neural network (FFNN) prediction (g), FFNN residuals (h). The same map symbology is applied to the prediction maps and residual maps respectively.

Mean annual temperature (MAT) is the only temperature-related covariate determined to be of high importance, and it has a non-linear impact on TSR (Fig. 4e). The annual range of temperature (ART) was the 3rd strongest predictor in Wang et al. (2011) and was the eighth important predictor in this analysis. When MAT is small (<5°C), it has a mostly negative impact on TSR, and the impact is stronger when MPDQ is greater (Fig. 4e). This indicates that environmental energy plays a limiting factor in areas where precipitation is relatively abundant. As MAT becomes larger, its impact on TSR becomes highly interactive with MPDQ and other covariates highlighting the energy-water dynamic impact on TSR (Francis and Currie, 2003). In a cross-scale study, Belmaker and Jetz (2010) did not observe Temperature/Energy impact on birds, mammals, and amphibians' richness at 20 km resolution, but this study confirms energy-water dynamics on tree species at the 20-km scale.

5. Discussion

This study examined the feasibility of using several machine learning algorithms to predict TSR in the continental U.S., and compared the performance of those algorithms to widely used GLMs. In this section, we discuss model performance, the ecological meaning of important predictor variables and their interactions with TSR in the continental U.S., and other modeling approaches for predicting species richness.

5.1. Model performance

Overall, all the machine learning models achieved R2 greater than 0.9 on the training dataset and R2 $\cong$ 0.9 on the test dataset, which suggests promising results for their use in predicting TSR. As far as we are aware, these accuracy levels have only been exceeded by one previous study predicting the species richness of angiosperms at a regional scale (Qian et al., 2015). Despite the presence of residual spatial autocorrelation (rSAC) in the RF and ANNs, their magnitudes were much smaller compared to the GLM, and they were more spatially dispersed (Fig. 5b, 4d, 4f, 4 h). Neither the RF nor ANNs were free from spatial autocorrelation of errors, but lower rSAC observed in both methods implies they predicted more reliable TSR spatial patterns compared to the GLM. This observed lower rSAC in RF and ANNs is presumably connected to the non-parametric nature of the machine learning methods, whereas GLM is sensitive to the requirement that observations are independent and identically distributed. Comparing within-model prediction accuracy, FFNN exhibited the smallest performance differences between training and test datasets, indicating that it potentially makes more accurate predictions on an independent, unseen dataset than RF and GRNN as large generalization errors of a model (i.e., high variance) typically indicate the model may overfit the training dataset. Thus, the FFNN model is recommended if transfer learning is applied to TSR prediction under future climatic scenarios or predictions of species richness of other organisms. However, the RF model performed similarly well with the added benefit of allowing a derivation of driver importance through the SHAP function with less computational intensity than the deep learning neural network models.

The spatial patterns of TSR in the continental United States predicted by the ANN and RF models (Fig. 5c, 5e and 5g) were very similar to the gridded FIA observations but more accurate than the GLM model (Fig. 5a) and more accurate than the GLM prediction by Wang et al (2011). For instance, our ANN and RF models predicted low and spatially heterogeneous patterns of TSR in the western U.S., while the GLM predicted patches of high TSR in the northwest (see high residuals in Fig. 5b); Wang et al. (2011) over-predicted TSR in much of the western United States. In the eastern U.S., both the results by Wang et al. (2011) and the GLM developed here show relatively high residuals. Overall, the RF and ANN models better captured the complex heterogeneous spatial patterns of TSR in the United States compared to the GLM model.

5.2. Modeled drivers of tree species richness

Both the GLM results and SHAP analysis of the RF indicate that forest area, precipitation, altitude, and temperature have the strongest impact on the geographic patterns of TSR in the continental United States. This finding aligns with previous studies at similar spatial resolutions (Wang et al., 2011; Kwon et al., 2018). While coefficients from the GLM model can indicate how TSR will change when the value of an environmental covariate changes, alone, they are insufficient for measuring the overall importance of a variable. This is because the value of GLM coefficients depends on the scale of the input predictor, and the scales of the variables used in this study vary considerably. In addition, the GLM model only reports a mean coefficient for each predictor. In comparison, the SHAP analysis of the RF model provides detailed explanations of how each predictor impacts a final predicted value as well as the interactions with other predictors.

According to the SHAP analysis, three precipitation-related covariates, aridity index (AI), mean annual precipitation (MPDQ), and mean annual precipitation (MAP) were among the top six predictors of TSR, with AI being the top predictor. This finding aligns with prior studies of TSR, which have found MAP to be a strong predictor of TSR in temperate ecoregions around the globe (Wang et al., 2011), and MPDQ to be a strong predictor of TSR in the eastern United States (Kwon et al., 2018). The fact that the impact of MPDQ on TSR was nearly the same in arid regions indicates that for western forests, decreasing precipitation during already dry seasons is not associated with intensifying negative TSR. This is concordant with the complex impacts of increasing drought on forest biodiversity in the United States (Clark et al., 2016) and suggests that monitoring should focus on compounding effects of climate change in the western United States. In addition, the non-linear and complex interactions between environmental energy variables, Energy-Water Dynamic (expressed as $PET\text{-}PET^2 + MAP$), and precipitation covariates, MAP and MPDQ highlight the importance of precipitation availability for TSR in the U.S and reveal the mixed impact of precipitation-related physiological stress and environmental energy. SHAP also uncovers that forest area has the strongest negative impact on TSR when forests are very small and this aligns with observational data in Little's range maps (e.g., Montoya et al., 2007) and emphases the importance of forest conservation efforts in low forested regions. Lastly SHAP reveals both altitude and temperature-related predictors have an important but more complex impact on TSR, which are compounded by other environmental covariates, such as precipitation and energy-water dynamics.

5.3. Capabilities for machine learning models to predict species richness

The results confirm that decision tree-based machine learning models, such as RF, detected more regional and subtle variations of TSR compared to GLM (França and Cabral, 2015; Li et al., 2017). RF has previously been criticized for producing lower prediction accuracy than GLM due to its inability to capture the non-symmetric error distribution of species richness (Lopatin et al., 2016). However, our results suggest that RF is a robust model for capturing TSR error distribution in the continental United States. This difference in findings may result from the fact that Lopatin et al. (2016) implemented a fixed RF model structure instead of training an RF model as suggested by machine learning approaches (Bzdok et al., 2017). Additionally, to our knowledge, this study is the first application of ANNs to TSR prediction using environmental covariates. The prediction accuracy exceeded other ANN applications of species richness estimation (Rocha et al., 2017; Franceschini et al., 2018) and suggest that FFNN is a valid model for TSR prediction.

Other machine learning approaches, such as a hybrid RF and kriging model (Li et al., 2017) have shown improved prediction accuracy of sponge species richness. We tested hybrid models for RF and FFNN with ordinary kriging to transform the TSR residuals, but we did not observe significant performance improvements: the hybrid RF model increased

R^2 marginally from 0.903 to 0.905 and lowed MAE from 2.90 to 2.85, while the R^2 and MAE of the hybrid FFNN models did not change (see Figure S3.1). The findings here suggest that this joint approach may depend on the chosen model, but more thorough integration of ANNs and RF and geostatistical tools are needed to fully assess the potential for species richness modeling. With the ease of access to cloud computing, state-of-the-art modeling software, and open access environmental and species databases, implementing machine learning algorithms for species richness modeling on large datasets is now feasible. For future studies, more advanced machine learning models, such as model averaging and deep learning, present an interesting development opportunity to improve accuracy.

6. Conclusions

Three machine learning models, RF and two ANN models (GRNN and FFNN), were employed to predict spatial patterns of TSR in the continental United States using plot-level tree species occurrence data from the FIA and 20 twenty environmental covariates. Results showed that all three machine learning models were powerful in predicting TSR with greater accuracy, less clustered residuals, and more reliable geographic patterns compared to the commonly used GLM. The FFNN model showed the highest prediction accuracy, the least generalization error, and a median TSR that was statistically similar to the observed TSR. However, the RF model had similar capabilities but with the added value that SHAP can be applied to explain the results with less computational intensity than the deep learning neural network models. The results of these models demonstrated that these machine learning models can be successfully applied to establish an accurate and reliable TSR prediction model. The modeling approach established in this study can potentially be transferred to the prediction of TSR using future climate scenarios or prediction of other species richness. SHAP analysis of the RF model revealed that the gridded TSR is best predicted by the aridity index, forest area, altitude, mean precipitation of the driest quarter, mean annual temperature, and mean annual precipitation. These covariates also show a non-linear relationship with TSR and interaction effects with other predictors. The findings suggest the importance of conservation efforts in preserving forest areas for tree species and further studying precipitation-related interacting factors to understand TSR patterns in the continental United States.

CRediT authorship contribution statement

Lian Brugere: Conceptualization, Methodology, Data curation, Software, Writing – original draft. **Youngsang Kwon:** Conceptualization, Methodology, Data curation, Supervision. **Amy E. Frazier:** Methodology, Writing – review & editing. **Peter Kedron:** Methodology, Writing – review & editing.

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.

Data availability

Data will be made available on request.

Acknowledgements

L. Brugere was supported by the Department of Earth Sciences at the University of Memphis (UofM). The authors appreciate the generous use of the High-Performance Computing Center and the computer lab at Center for Applied Earth Science and Engineering Research of the UofM. A. Frazier is supported by United States National Science Foundation grant #2225076.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S. et al. 2015. TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from https://tensorflow.org/.

Agarap, A.F. 2018. Deep Learning using Rectified Linear Units (ReLU). https://arxiv.org/abs/1803.08375.

Bechtold, W.A., & Patterson, P.L. 2005. The Enhanced Forest Inventory and Analysis Program: National Sampling Design and Estimation Procedures. Vol. 80. US Department of Agriculture Forest Service, Southern Research Station: Asheville, NC, USA.

Belmaker, J., Jetz, W., 2010. Cross-scale variation in species richness–environment associations. Glob. Ecol. Biogeogr. 20 (3), 464–474. https://doi.org/10.1111/j.1466-8238.2010.00615.x.

Bergstra, J., Komer, B., Eliasmith, C., Yamins, D., Cox, D., 2015. Hyperopt: A Python library for model selection and hyperparameter optimization. Comput. Sci. Discov. 8 https://doi.org/10.1088/1749-4699/8/1/014008.

Bland, L.M., Collen, B., Orme, C.D.L., Bielby, J., 2015. Predicting the conservation status of data-deficient species. Conserv. Biol. 29 (1), 250–259. https://doi.org/10.1111/cobi.12372.

Breiman, L., 2001. Random forest. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Bzdok, D., Krzywinski, M., Altman, N., 2017. Machine learning: A primer. Nat. Methods 14, 1119–1120. https://doi.org/10.1038/nmeth.4526.

Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P., Narwani, A., Mace, G.M., Tilman, D., Wardle, D.A., et al., 2012. Biodiversity loss and its impact on humanity. Nature 486, 59–67. https://doi.org/10.1038/nature11148.

Chollet, F., et al., 2015. Keras. Retrieved from, GitHub https://github.com/fchollet/keras.

Christin, S., Hervet, É., Lecomte, N., 2019. Applications for Deep Learning in Ecology. Methods Ecol. Evol. 10 (10), 1632–1644. https://doi.org/10.1111/2041-210X.13256.

Clark, J.S., et al., 2016. The impacts of increasing drought on forest dynamics, structure, and biodiversity in the United States. Glob. Chang. Biol. 22 (7), 2329–2352.

Cleland, E.E., 2011. Biodiversity and Ecosystem Stability. Nature Education Knowledge 3 (10), 14.

Fan, W., & Waring, R.H. 2009. Actual Evapotranspiration (AET) and tree species richness in the eastern U.S.A. In: McWilliams, W., Moisen, G., Czaplewski, R., comps. Forest Inventory and Analysis (FIA) Symposium 2008; October 21-23, 2008; Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 13 p.

França, S., Cabral, H.N., 2015. Predicting fish species richness in estuaries: Which modelling technique to use? Environ. Model. Softw. 66, 17–26. https://doi.org/10.1016/j.envsoft.2014.12.010.

Franceschini, S., Gandola, E., Martinoli, M., Tancioni, L., Scardi, M., 2018. Cascaded neural networks improving fish species prediction accuracy: the role of the biotic information. Sci. Rep. 8, 4581. https://doi.org/10.1038/s41598-018-22761-4.

Francis, A.P., Currie, D.J., 2003. A globally consistent richness-climate relationship for angiosperms. Am. Nat. 161, 523–536. https://doi.org/10.1086/368223.

Gentry, A.H., 1988. Changes in plant community diversity and floristic composition on environmental and geographical gradients. Ann. Mo. Bot. Gard. 75 (1), 1–34. https://doi.org/10.2307/2399464.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press, Cambridge, MA.

Haghverdi, K., Kooch, Y., 2019. Effects of diversity of tree species on nutrient cycling and soil-related processes. Catena 178, 335–344. https://doi.org/10.1016/j.catena.2019.03.041.

Huang, Y., Chen, Y., Castro-Izaguirre, N., Baruffol, M., Brezzi, M., Lang, A., Li, Y., Härdtle, W., von Oheimb, G., Yang, X., et al., 2018. Impacts of species richness on productivity in a large-scale subtropical forest experiment. Science 362 (6410), 80–83. https://doi.org/10.1126/science.aat6405.

Hubbell, S.P., 2001. The unified neutral theory of biodiversity and biogeography. Princeton University Press, Princeton, New Jersey, USA http://www.jstor.org/stable/j.ctt7rj8w.

Iverson, L.R., Prasad, A.M., 2001. Potential changes in tree species richness and forest community types following climate change. Ecosystems 4, 186–199. https://doi.org/10.1007/s10021-001-0003-6.

Keil, P., Chase, J.M., 2019. Global patterns and drivers of tree diversity integrated across a continuum of spatial grains. Nat. Ecol. Evol. 3, 390–399. https://doi.org/10.1038/s41559-019-0799-0.

Kingma, D.P. & Ba, J.L. 2015. Adam: A Method for Stochastic Optimization. https://arxiv.org/abs/1412.6980.

Kwon, Y., Larsen, C., Lee, M., 2018. Tree species richness predicted using a spatial environmental model including forest area and frost frequency, eastern USA. PLoS One 13 (9), e0203881.

Kwon, Y., Lee, T., Lang, A., Burnette, D., 2019. Assessment on latitudinal tree species richness using environmental factors in the southeastern United States. PeerJ 7, e6781.

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J., Aulagnier, S., 1996. Application of neural networks to modelling nonlinear relationships in ecology. Ecol. Model. 90 (1), 39–52. https://doi.org/10.1016/0304-3800(95)00142-5.

Li, J., Alvarez, B., Siwabessy, J., Tran, M., Huang, Z., Przeslawski, R., Radke, L., Howard, F., Nichol, S., 2017. Application of random foresolarcticdsed linear model

and their hybrid methods with geostatistical techniques to count data: Predicting sponge species richness. Environ. Model. Softw. 97, 112–129. https://doi.org/10.1016/j.envsoft.2017.07.016.
Liaw, A., Wiener, M., 2001. Classification and Regression by randomForest. R News 2 (3), 18–22.
Linardatos, P., Papastefanopoulos, V.; Kotsiantis, S. 2021. Explainable AI: A Review of Machine Learning Interpretability Methods. Entropy 2021, 23, 18. https://dx.doi.org/10.3390/e23010018.
Liu, Z., Peng, C., Work, T., Candau, J., DesRochers, A., Kneeshaw, D., 2018. Application of machine-learning methods in forest ecology: recent progress and future challenges. Environ. Rev. 26 (10), 339–350. https://doi.org/10.1139/er-2018-0034.
Lopatin, J., Dolos, K., Hernández, J., Galleguillos, M., Fassnacht, F.E., 2016. Comparing Generalized Linear Models and random forest to model vascular plant species richness using LiDAR data in a natural forest in central Chile. Remote Sens. Environ. 173 https://doi.org/10.1016/j.rse.2015.11.029.
Lundberg, S.M.; Lee, S.I. 2017. A unified approach to interpreting model predictions. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA. pp. 4765–4774.
MacArthur, R.H., Diamond, J.M., Karr, J.R., 1972. Density Compensation in Island Faunas. Ecology 53 (2), 330–342. https://doi.org/10.2307/1934090.
MacArthur, R.H., Wilson, E.O., 1967. The theory of island biogeography. Princeton University Press, Princeton, N.J.
Maina, J.M., 2021. Identifying global and local drivers of change in mangrove cover and the implications for management. Glob. Ecol. Biogeogr. 30 (10), 2057–2069. https://doi.org/10.1111/geb.13368.
McRoberts, R.E., Bechtold, W.A., Patterson, P.L., Scott, C.T., Reams, G.A., 2005. The enhanced Forest Inventory and Analysis program of the USDA Forest Service: Historical perspective and announcement of statistical documentation. J. For. 3 (6), 304–308.
Molnar, C. 2020. Interpretable machine learning. https://christophm.github.io/interpretable-ml-book/.
Montoya, D., Rodríguez, M.A., Zavala, M.A., Hawkins, B.A., 2007. Contemporary richness of holarctic trees and the historical pattern of glacial retreat. Ecography 30, 173–182. https://doi.org/10.1111/j.2006.0906-7590.04873.x.
Ouyang, S. Xiang, W, Gou, M., et al. 2020. Global Ecology and Biogeography, 30(2):500-513. https://doi.org/10.1111/geb.13235.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., Louppe, G., 2011. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12, 2825–2830.
Pereira, H.M., Ferrier, S., Walters, M., et al., 2013. Essential biodiversity variables. Science 339 (6117), 277–278. https://doi.org/10.1126/science.1229931.
Qian, H., Wien, J.J., Zhang, J. & Zhang, Y. 2015. Evolutionary and ecological causes of species richness patterns in North American angiosperm trees. Ecography, 38, 241–250. https://doi.org//10.1111/ecog.00952. https://doi.org/10.1038/s41586-019-0912-1.
Rocha, J., Peres, C., Buzzo, J., Souza, V., Krause, E., Bispo, P., Frei, F., Costa, L., Branco, C., 2017. Modeling the species richness and abundance of lotic macroalgae based on habitat characteristics by artificial neural networks: a potentially useful tool for stream biomonitoring programs. J. Appl. Phycol. 29 https://doi.org/10.1007/s10811-017-1107-5.
Rosenzweig, M.L., 1995. Species diversity in space and time. Cambridge University Press, Cambridge, MA, p. 436.
Sarr, D., Hibbs, D., Huston, M., 2005. A Hierarchical Perspective of Plant Diversity. Q. Rev. Biol. 80 (2), 187–212. https://doi.org/10.1086/433058.
Schuldt, A., Fornoff, F., Bruelheide, H., Klein, A.-M., Staab, M., 2017. Tree species richness attenuates the positive relationship between mutualistic ant-hemipteran interactions and leaf chewer herbivory. Proc. R. Soc. B 284 (862). https://doi.org/10.1098/rspb.2017.1489.
Seabold, S. & Perktold, J. 2010. Statsmodels: Econometric and Statistical Modeling with Python. Proceedings of the 9th Python in Science Conference. 2010.
Shevchuk, Y. 2015. NeuPy: Neural Networks in Python. http://neupy.com/pages/home.html.
Shapley, L.S., 1951. *Notes on the N-person Game.* Rand Corporation.
Specht, D., 1991. A General Regression Neural Network. IEEE Trans. Neural Netw. 2 (6), 568–578. https://doi.org/10.1109/72.97934.
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929–1958.
Svenning, J.-C., Skov, F., 2007. Could the tree diversity pattern in Europe be generated by postglacial dispersal limitation? Ecol. Lett. 10 (6), 453–460. https://doi.org/10.1111/j.1461-0248.2007.01038.x.
Swenson, N.G., et al., 2011. The biogeography and filtering of woody plant functional diversity in North and South America. Glob. Ecol. Biogeogr. 21 (8), 798–808. https://doi.org/10.1111/j.1466-8238.2011.00727.x.
Tilman, D. 1982. Resource Competition and Community Structure. MPB-17, Volume 17. Princeton, New Jersey: Princeton University Press. https://doi.org/10.2307/j.ctvx5wb72.
Vetaas, O.R., Grytnes, J.-A., 2002. Distribution of vascular plant species richness and endemic richness along the Himalayan elevation gradient in Nepal. Glob. Ecol. Biogeogr. 11 (4), 291–301. https://doi.org/10.1046/j.1466-822X.2002.00297.x.
Wang, Z., Fang, J., Tang, Z., Lin, X., 2011. Patterns, determinants and models of woody plant diversity in China. Proc. R. Soc. B 278 (1715). https://doi.org/10.1098/rspb.2010.1897.
Wasserstein, R.L., Lazar, N.A., 2016. The ASA's statement on *p*-values: context, process, and purpose. Am. Stat. 70, 129–133. https://doi.org/10.1080/00031305.2016.1154108.
Welchowski, T., Maloney, K.O., Mitchell, R., et al., 2021. Techniques to Improve Ecological Interpretability of Black-Box Machine Learning Models. J. Agric. Biol. Environ. Stat. https://doi.org/10.1007/s13253-021-00479-7.
Woodall, C.W., Oswalt, C.M., Westfall, J.A., Perry, C.H., Nelson, M.D., Finley, A.O., 2010. Selecting tree species for testing climate change migration hypotheses using forest inventory data. For. Ecol. Manage. 259 (4), 778–785.
Wu, J., Liang, S., 2018. Developing an Integrated Remote Sensing Based Biodiversity Index for Predicting Animal Species Richness. Remote Sens. (Basel) 10 (5), 739. https://doi.org/10.3390/rs10050739.
Zhu, K., Woodall, C., Monteiro, J., Clark, J., 2015. Prevalence and strength of density dependent tree recruitment. Ecology 96 (9), 2319–2327. https://doi.org/10.1890/14-1780.1.