Spatial mapping of key plant functional traits in terrestrial ecosystems across China

Nannan An¹,²,³, Nan Lu²,³, Weiliang Chen², Yongzhe Chen²,⁴, Hao Shi²,³,
Fuzhong Wu¹, Bojie Fu²,³

¹Key Laboratory for Humid Subtropical Eco-Geographical Processes of the Ministry of Education, School of Geographical Sciences, Fujian Normal University, Fuzhou 350117, PR China
²State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences (CAS), Beijing 100085, PR China
³University of Chinese Acedemy of Sciences, Beijing 101408, PR China
⁴Department of Geography, The University of Hong Kong, Hongkong, 999077 , PR China

Correspondence to: Nan Lu (nanlv@rcees.ac.cn)
Abstract

Trait-based approaches are of increasing concern in predicting vegetation changes and linking ecosystem structures to functions at large scales. However, a critical challenge for such approaches is acquiring spatially continuous plant functional trait maps. Here, six key plant functional traits were selected as they can reflect plant resource acquisition strategies and ecosystem functions, including specific leaf area (SLA), leaf dry matter content (LDMC), leaf N concentration (LNC), leaf P concentration (LPC), leaf area (LA) and wood density (WD). A total of 34589 in-situ trait measurements of 3447 seed plant species were collected from 1430 sampling sites in China and were used to generate spatial plant functional trait maps (~1 km), together with environmental variables and vegetation indices based on two machine learning models (random forest and boosted regression trees). To obtain the optimal estimates, a weighted average algorithm was further applied to merge the predictions of the two models to derive the final spatial plant functional trait maps. The models showed a good accuracy in estimating WD, LPC and SLA, with average $R^2$ values ranging from 0.48 to 0.68. In contrast, both the models had weak performance in estimating LDMC, with average $R^2$ values less than 0.30. Meanwhile, LA showed considerable differences between two models in some regions. Climatic effects were more important than those of edaphic factors in predicting the spatial distribution of plant functional traits. Estimates of plant functional traits in northeast China and the Qinghai-Tibet Plateau had relatively high uncertainties due to sparse samplings, implying a need of more observations in these regions in future. Our spatial trait maps could provide critical supports for trait-based vegetation models and allow exploration into the relationships between vegetation characteristics and ecosystem functions at large scales. The six plant functional traits maps for China with 1 km spatial resolution are now available at https://figshare.com/s/c527c12d310cb8156ed2 (An et al., 2023).
1 Introduction

Climate change has been affecting vegetation distributions and biogeochemical cycling globally and altering their feedbacks to climate system (Kirilenko et al., 2000; Finzi et al., 2011; Jónsdóttir et al., 2022). Dynamic global vegetation models (DGVMs) are powerful tools for predicting changes in vegetation and ecosystem-atmosphere exchanges (e.g., water, carbon and nutrient cycling) in a changing climate (Foley et al., 1996; Peng, 2000). However, conventional DGVMs are still insufficient realistic, largely due to their dependence on the plant functional types (PFTs) assumption (Sitch et al., 2008; Yurova and Volodin, 2011; Scheiter et al., 2013). PFTs in conventional DGVMs commonly have fixed attributes (mostly trait values) (Van Bodegom et al., 2012; Wullschleger et al., 2014) that do not reflect plant adaptation to environments, limiting the quantification of carbon-water-nutrient feedback between terrestrial ecosystems and the atmosphere (Zaehle and Friend, 2010; Liu and Yin, 2013). Trait-based approaches can provide robust theoretical basis for developing the next generation of DGVMs (Van Bodegom et al., 2012; Sakschewski et al., 2015; Matheny et al., 2017). Plant functional traits, which are closely associated with ecosystem functions (Diaz et al., 2004; Yan et al., 2023), can effectively reflect response and adaptation of plants to environmental conditions (Myers-Smith et al., 2019; Qiao et al., 2023).

Attempts to predict spatially continuous trait maps have been conducted at regional to global scales (Madani et al., 2018; Moreno-Martínez et al., 2018; Boonman et al., 2020; Loozen et al., 2020; Dong et al., 2023). Webb et al. (2010) proposed that the environment creates a filtered trait distribution along an environmental gradient, and such trait-environment relationships offer fundamental supports to predict the spatial distribution of plant functional traits through extrapolating local trait measurements. Boonman et al. (2020) mapped the global patterns of specific leaf area (SLA), leaf N concentration (LNC) and wood density (WD) based on a set of climate and soil variables. As the number of available regional and global trait databases increases (Wang et al., 2018; Kattge et al., 2020), trait-environment relationships are becoming increasingly quantitative and accurate (Bruelheide et al., 2018; Myers-Smith et al., 2019). Alternatively, remote sensing approaches, such as empirical methods and physical radiative transfer models (e.g., partial least squares regression and PROSPECT model), have been developed to estimate plant physiological, morphological and chemical traits (e.g., leaf chlorophyll content, SLA, LNC and leaf dry matter content (LDMC)) (Darvishzadeh et al., 2008; Romero et al., 2012; Ali et al., 2016). Vegetation indices, such as normalized difference vegetation index (NDVI) and enhanced vegetation index (EVI), have been successful in estimating plant functional traits of crops, grasslands and forests (Clevers and Gitelson, 2013; Li et al., 2018; Loozen et al., 2018). Loozen et al. (2020) demonstrated that EVI was the most important predictor for mapping the spatial pattern of canopy nitrogen in European forests. Admittedly, a recent study has suggested that combining environmental variables and vegetation indices can improve the predictive accuracy of canopy
nitrogen compared to those based on vegetation indices alone (Loozen et al., 2020).

Although there have been reports on plant functional trait distribution in China in some global or regional researches (Yang et al., 2016; Butler et al., 2017; Madani et al., 2018; Moreno-Martínez et al., 2018; Boonman et al., 2020), they are still large uncertainties in characterizing the spatial distribution of plant functional traits in China. First, global studies generally have relatively few, unevenly distributed sampling sites across China (Butler et al., 2017; Madani et al., 2018; Boonman et al., 2020), impeding our understanding of the true spatial characteristics of trait variability. Second, the spatial pattern of traits among these studies are usually inconsistent. For example, Moreno-Martínez et al. (2018) and Madani et al. (2018) demonstrated that SLA values were low in the southeast areas but high in the southwest areas of China, whereas Boonman et al. (2020) found the opposite. Third, most studies focused on leaf traits (Yang et al., 2016; Loozen et al., 2018; Moreno-Martínez et al., 2018), whereas traits associated with the whole-plant strategies, such as WD, were ignored. Therefore, mapping and verifying the spatial patterns of key functional traits that reflect the whole plant economics spectrum in China is a top priority.

In this study, our main objective was to generate spatial maps for several key plant functional traits, through combining field measurements, environmental variables and vegetation indices. We selected six plant functional traits including SLA, LDMC, LNC, LPC, LA and WD. As key leaf economics traits, SLA, LDMC, LNC and LPC were selected because they are closely linked to plant growth rate, resource acquisition and ecosystem functions (Wright et al., 2004; Diaz et al., 2016). LA is indicative of the trade-off between carbon assimilation and water-use efficiency (Wright et al., 2017), and WD reflects the trade-off between plant growth rate and support cost, with a higher WD linked to a lower growth rate, a higher survival rate and a higher biomass support cost (King et al., 2006). For each plant functional trait, we predicted spatial patterns at a 1 km resolution using an ensemble modelling algorithm based on two machine learning methods (i.e., random forest and boosted regression trees).

2 Materials and Methods

2.1 Overview

The spatial maps of plant functional traits in China were generated based on machine learning methods trained by a large dataset of in-situ field measurements, environmental variables and vegetation indices in three steps (Fig. 1). First, in-situ field measurements of six plant functional traits were collected from TRY and China databases as well as published literature, and the PFTs of plant species were classified based on plant growth form, leaf type and leaf phenology. Multiple gridded predictors of climate, soil, topography and vegetation indices were used after avoiding the collinearity among them. Second, random forest and boosted regression trees were used to train the relationships between plant functional traits and predictors for each PFT individually. Third, the spatial abundance of each PFT within 1 km grid cell was calculated using land cover map (100...
m). Community-weighted trait values within 1 km grid cell were calculated based on these abundances of each PFT and their predicted trait values in Step 2. To reduce the variability of different single-models, we derived the final spatial maps of plant functional traits using ensemble model algorithm to merge the predictions of random forest and boosted regression trees according to their cross-validated $R^2$ values.

Figure 1. Methodological workflow for spatial mapping of plant functional traits. Trait mapping is performed in three steps. Step 1: in-situ field measurement of plant functional traits, PFT classification of plant species and gridded predictors were collected. Step 2: two machine learning methods were used to predict trait values by training the field measurements and predictors for each PFT. Step 3: spatialization of trait maps by calculating the abundance of each PFT using 100 m land cover map and predicted trait values within 1 km grid cells. PFT, plant functional type; RF, random forest; BRT, boosted regression trees.

2.2 Plant functional trait collection and data processing

The information on the six plant functional traits and their ecological meanings are described in Table 1. Plant trait data was obtained and collected via two main sources. The first source was public trait databases, including the TRY database (Kattge et al., 2020) and the China Plant Trait Database (Wang et al., 2018). The second source was from literature (listed in Appendix A). To ensure data quality and comparability, we only included trait observations that met the following five criteria: 1) Measurements must be obtained from natural terrestrial fields in order to minimize the influences of management disturbance, and observations from cropland, aquatic habitat, control experiments and gardens were excluded; 2) According to the mass ratio hypothesis, the effect of plant species on ecosystem functioning is determined to an overwhelming extent by the traits and functional diversity of the dominant species and is relatively insensitive to the richness of subordinate species (Grime, 1998). Thus, we only included studies that measured plant trait
observations from all species or dominant species within a community; 3) In order to consider the intraspecific trait variation, when the same species occurred in the same sampling site from different studies, we included all original observed data from different studies rather than averaging the values at the species level (Jung et al., 2010; Siefert et al., 2015); 4) Plant trait observations must be made on mature and healthy plant individuals, so some specific growth stages (e.g., seedling) and size classes (e.g., sapling) were excluded to reduce the confounding effect of ontogeny and seasonality (Thomas, 2010); 5) We only included studies with clear geographical coordinates to match predictor variables. The sampling location and sampling time information were also included in the dataset. The sampling time mostly focused on the growing season of a year (i.e., May-October), which ensures the relative consistency of sampling time to minimize the effects of seasonality. Plant functional traits must be sampled and measured according to standardized measurement procedures (Perez-Harguindeguy et al., 2013) to reduce the variation and uncertainty among different data sources. In this study, we included SLA measurements on sun-leaves, and WD measurements on main stem of woody species.

Table 1 Description of plant functional traits selected in this study and their relevant ecosystem functions.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Abbreviation</th>
<th>Description</th>
<th>Relevant ecosystem functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specific leaf area</td>
<td>SLA</td>
<td>As a core leaf economics trait (Wright et al., 2004), it is related to trade-off between leaf lifespan and C acquisition as well as light competition (Reich et al., 1991)</td>
<td>Productivity, litter decomposition, competitive ability (Bakker et al., 2011; Smart et al., 2017)</td>
</tr>
<tr>
<td>Leaf dry matter content</td>
<td>LDMC</td>
<td>Strongly related to resource availability and potential growth rate (Hodgson et al., 2011)</td>
<td>Productivity, litter decomposition, herbivore resistance, and drought tolerance (Bakker et al., 2011; Smart et al., 2017; Blumenthal et al., 2020)</td>
</tr>
<tr>
<td>Leaf N concentration</td>
<td>LNC</td>
<td>As a core leaf economics trait, it is strongly related to photosynthetic capacity (Wright et al., 2004)</td>
<td>Productivity, nutrient cycling, litter decomposition (LeBauer and Treseder, 2008; Bakker et al., 2011)</td>
</tr>
<tr>
<td>Leaf P concentration</td>
<td>LPC</td>
<td>As a core leaf economics trait, it is strongly related to photosynthetic capacity (Wright et al., 2004)</td>
<td>Productivity, nutrient cycling, litter decomposition (LeBauer and Treseder, 2008; Bakker et al., 2011)</td>
</tr>
<tr>
<td>Leaf area</td>
<td>LA</td>
<td>Trade-off between C assimilation and water use efficiency, it is related to energy balance (Wright et al., 2017)</td>
<td>Productivity (Li et al., 2020)</td>
</tr>
<tr>
<td>Wood density</td>
<td>WD</td>
<td>A measure of carbon investment, representing the trade-off between growth and mechanical support (Martínez-Vilalta et al., 2010)</td>
<td>Drought tolerance, productivity (Hoeber et al., 2014; Liang et al., 2021)</td>
</tr>
</tbody>
</table>

The plant trait data was checked for possible errors and corrected in three steps as follows. First, species name and taxonomic nomenclature were corrected and standardized according to the Plant List (http://www.theplantlist.org/) using the “plantlist” package. Second, illogical values, repeated values and outliers were removed, which were defined by observations exceeding 1.5
standard deviations of the mean trait value for a given species (Kattge et al., 2011). Third, we appended information on plant growth form, leaf type and leaf phenology from the TRY categorical traits database (https://www.try-db.org/TryWeb/Data.php3) and Flora Reipublicae Popularis Sinicae (http://www.iplant.cn/frps), which were used to match species names to PFTs. We associated each species with a corresponding PFT based on plant growth form (tree, shrub and grass), leaf type (broadleaf and needleleaf) and leaf phenology (evergreen and deciduous). For example, the information on Salix matsudana is: tree, deciduous and broadleaf, thus, we were able to associate the PFT of deciduous broadleaf forest (DBF) to this species. The species that did not correspond to any PFT were discarded. After these treatments, we collected a total of 34589 trait measurements from 1430 sampling sites for our database, representing 3447 species from 195 families and 1066 genera (Fig. 2). Information on the statistics for the six plant functional traits collected in this study is shown in Table B1 in Appendix B.

![Figure 2. Location distribution and land cover map in China.](image)

### 2.3 Preparing predictor variables

#### 2.3.1 Climate data

Twenty-one climate variables were used in this study, including 19 bioclimate variables, solar radiation (RAD) and aridity index (AI) (Table B2 in Appendix B). The 19 bioclimate variables and RAD were obtained from the WorldClim version 2.1 for the period from 1970 to 2000 (https://www.worldclim.org/data/worldclim21.html). The AI data was extracted from the CGIAR Consortium of Spatial Information (CGIAR-CSI) website for the period from 1970 to 2000 (http://www.csi.cgiar.org) (Trabucco and Zomer, 2018). The spatial resolution of climate data is 1 km.

#### 2.3.2 Soil data

Twelve soil variables were included in this study, representing the different aspects of soil properties, i.e. soil texture, bulk density (BD), pH and soil nutrients (Table B2 in Appendix B). All
soil variables were extracted from the Soil Database of China for Land Surface Modeling (http://globalchange.bnu.edu.cn/research/soil2) (Shangguan et al., 2013). Given the importance of topsoil properties on community composition (Bohner, 2005), we averaged the first four layers to represent the topsoil properties (~ 30 cm) in our study. The spatial resolution is 1 km.

2.3.3 Topography
The topographic variable was elevation. Elevation data was extracted from the STRM 90m dataset in China, based on the SRTM V4.1 database (https://www.resdc.cn/data.aspx?DATAID=123). The spatial resolution is 1 km.

Given the collinearity among climate and soil variables, we reduced the dimensionality of these predictors based on Pearson’s correlation coefficient (r) (Figs. B1 and B2 in Appendix B). Among a set of highly correlated variables (r > 0.75), only one variable was retained in subsequent analysis to ensure a combination of different environmental variables. The final selection of environment predictors included nineteen variables: mean annual temperature (MAT), mean diurnal range (MDR), min temperature of coldest quarter (Tmin), max temperature of warmest quarter (Tmax), temperature seasonality (TS), mean annual precipitation (MAP), precipitation seasonality (PS), precipitation of wettest quarter (PEQ), precipitation of driest quarter (PDQ), AI, RAD, elevation, soil sand content (SAND), pH, BD, soil total N (STN), soil total P (STP), soil available P (SAP), soil alkali-hydrolysable N (SAN) and cation exchange capacity (CEC).

2.3.4 Vegetation indices
Three categories of vegetation indices were included in this study (Table B2 in Appendix B). First, EVI was extracted from the MOD13A3 V006 product (https://lpdaac.usgs.gov/products/mod13a3v006/). This product is available as a monthly average with spatial resolution of 1 km, ranging from January 2000 to December 2018. Second, MODIS reflectance data was also extracted from the MOD13A3 V006 product, including MIR reflectance, NIR reflectance, red reflectance and blue reflectance. Third, the MERIS terrestrial chlorophyll index (MTCI) was extracted from the Natural Environment Research Council Earth Observation Data Centre (NERC-NEODC, 2005) (https://data.ceda.ac.uk/). MTCI data is available globally as a monthly average at 4.63 km spatial resolution, and ranges from June 2002 to December 2011. It is noted that valid MTCI values should be greater than 1, so our study deleted any values less than 1.

To avoid collinearity, we also reduced the dimensionality of vegetation indices based on Pearson’s correlation coefficient (r) (Fig. B3 in Appendix B). Most selected variables were related to growing seasons due that plant functional traits were measured during the growing season. Furthermore, based on the results of Pearson’s correlation coefficient (r), MTCI, MIR, NIR, red and blue in January showed low correlations with those in growing season, thus they were included in subsequent analysis. The final selection included 36 variables: annual EVI, EVI (May, June, July, August and September), MTCI, MIR, NIR, red and blue (all for January, June, July, August and September).
Both environmental variables and vegetation indices variables were resampled to a consistent spatial resolution of 1 km using the nearest neighborhood method.

PFT is also an important factor in influencing the variation of plant functional traits (Verheijen et al., 2016; Loozen et al., 2020), thus the trait predictions were performed for each PFT individually. We used the 2015 land cover map at a 100 m spatial resolution to calculate the relative abundance of each PFT within 1 km grid cells, which was extracted from the Copernicus Global Land Service (CGLS-LC100, Version 3) (https://land.copernicus.eu/global/products/lc) (Buchhorn et al., 2020). We focused on natural terrestrial vegetation, so all artificial or crop areas were thus eliminated in our dataset. Seven categories were included: evergreen needleleaf forest (ENF), evergreen broadleaf forest (EBF), deciduous needleleaf forest (DNF), deciduous broadleaf forest (DBF), shrubland (SHL), grassland (GRL) and bare/sparse vegetation.

2.4 Model fitting and validation

To predict spatial patterns of plant functional traits, we used two machine learning models, i.e., random forest and boosted regression trees.

Random forest is an ensemble machine learning method based on classification and regression trees using collections of regression trees to classify observations according to a set of predictive variables (Breiman, 2001). This method repeatedly constructs a set of trees from random samples of training data, and the final prediction is produced by integrating the results of all individual trees, which makes it a robust method. The model is controlled by two main parameters: the number of sampled variables (mtry) and the number of trees (ntree). The mtry was set to range from 1 to 57 (at an interval of 1), and the ntree was set as 500, 1000, 2000, 5000 and 10000 in subsequent runs. This analysis was performed using the ‘randomForest’ function in the ‘randomForest’ package (Liaw and Wiener, 2002).

Boosted regression trees are machine learning methods based on generalized boosted regression models and using a boosting algorithm to combine many sample tree models to optimize predictive performance (Elith et al., 2006). There is no need for prior data transformation or the elimination of outliers, and this method can fit complex non-linear relationships while automatically handling interaction effects between predictors (Elith et al., 2008). The four parameters to optimize in these models are the number of trees, interaction depth, learning rate and bag fractions. We varied the parameter settings to find the optimal parameter combination that achieves minimum predictive error. The number of trees was set to 3000, the interaction depth varied from 1 to 7 (at an interval of 1), the learning rate was set to 0.001, 0.01, 0.05 and 0.1, and the bag fraction was set to 0.5, 0.6, 0.7 and 0.75. PFT was used as a dummy variable in the boosted regression trees models. This analysis was conducted using the ‘gbm’ function in the ‘gbm’ package (Ridgeway, 2006).

We built separate predictive model for each plant functional trait. To select the optimal parameter combination and to evaluate the final model performance for each trait, we calibrated
the models 10 times using randomly selected 80% of the data for training the models and validating against the remaining 20% based on cross-validation (Table B3 in Appendix B). The predictive performance was evaluated by regressing the predicted and observed trait values from all repetitions of the cross-validation. The fitting performances of the random forest and boosted regression trees were evaluated using determinate coefficient ($R^2$), normalized root-mean-square error (NRMSE) and mean absolute error (MAE). These scores are calculated following Eq. (1), Eq. (2) and Eq. (3):

\[ R^2 = 1 - \frac{\sum_{i=1}^{n}(p_i - o_i)^2}{\sum_{i=1}^{n}(p_i - \bar{o}_i)^2} \]

\[ NRMSE = \sqrt{\frac{\frac{1}{n} \sum_{i=1}^{n}(p_i - o_i)^2}{p_{max} - p_{min}}} \]

\[ MAE = \frac{1}{n} \sum_{i=1}^{n} |o_i - p_i| \]

where $p_i$ and $o_i$ are the predictive values and observed values, respectively; $\bar{o}_i$ is the mean of the observed values.

To quantify the relative importance of each predictor across the two models consistently, we used the method proposed by Thuiller et al. (2009). This method applies correlation between the standard predictions fitted with the original data and predictions where the variable under investigation has been randomly permuted. If the correlation is high, which indicates little difference between the two predictions, the variable permuted is considered not important for the model. This step was repeated multiple times for each predictor, and the mean correlation coefficient over runs was recorded. Then the relative importance of each predictor was quantified as one minus the Spearman rank correlation coefficient (see Boonman et al., 2020). In addition, we used generalized additive models to fit the relationships between plant functional traits and the most important variables using the ‘gam’ function in the ‘mgcv’ package.

### 2.5 Generation of plant functional trait maps and model performance

The generation of spatial maps of plant functional was performed in three steps. First, we predicted trait values for each natural PFT (e.g., EBF, ENF, DBF, DNF, SHL and GRL) within 1 km grid cell separately. Second, the abundance of individual natural PFT within 1 km grid cell was estimated using a land cover map with a spatial resolution of 100 m. Third, refer to the Eq. (4) that has been widely applied in a community (Garnier et al., 2004), the final trait value in a given 1 km grid cell was calculated as the sum of the predicted trait values multiplying by corresponding abundance of each natural PFT.

\[ CWM = \sum_{i=1}^{n} W_i X_i \]

where $n$ is the total number of PFT in a given grid; $W_i$ is the relative abundance of the $i$th natural PFT; $X_i$ is the predicted trait value of the $i$th natural PFT.

To reduce the variability of different single-models and to construct a more stable and accurate model, the ensemble model was further applied to merge the predictions of random forest
and boosted regression trees according to their cross-validated $R^2$ values. The predictive value of ensemble model was calculated in a given grid cell as described by Eq. (5) (Marmion et al., 2009). The model accuracy was calculated by regressing the predictive values of ensemble model against the observed trait values.

$$Pred_{EM_t} = \frac{\sum_{m=1}^{5}(pred_{m,t} \times r_{m,t}^2)}{\sum_{m=1}^{5} r_{m,t}^2}$$  \hspace{1cm} (5)$$

where $Pred_{EM_t}$ is the predictive values of $t$ trait in the ensemble model; $pred_{m,t}$ is the predictive values of $t$ trait in $m$ model; $r_{m,t}^2$ is the cross-validated $R^2$ of $t$ trait in $m$ model.

To evaluate the model performance (i.e. the variability in the prediction across models), the coefficient of variation (CV) was calculated as the difference between the predictions of random forest and boosted regression trees methods and the ensemble prediction. CV is calculated as following Eq. (6):

$$CV_t = \frac{\sqrt{\sum_{m=1}^{5}(pred_{m,t} - obs_t)^2}}{\sum_{m=1}^{5} r_{m,t}^2}$$

$$obs_t = \frac{\sum_{m=1}^{5} r_{m,t}^2}{5}$$  \hspace{1cm} (6)$$

where $pred_{m,t}$ is the predictive values of $t$ trait in $m$ model; $obs_t$ is the values of $t$ trait in the ensemble model; $r_{m,t}^2$ is the cross-validated $R^2$ of $t$ trait in $m$ model.

### 2.6 Uncertainty assessments

Multivariate environmental similarity surface analysis (MESS) was used to identify the range of the extrapolated predictor values across the locations in the plant trait dataset (Elith et al., 2010). This method is often used to evaluate the extent of extrapolation and the applicability domain. If the values are negative, this indicates that at a given grid cell, at least one predictor variable is outside the extent of referenced predictor layer. This analysis was conducted using the ‘mess’ function in the ‘dismo’ package.

All analyses were performed in R 4.0.2 (R Core Team, 2020).

### 3 Results

#### 3.1 Performances of prediction models

Cross-validation showed that the performance of the predictive models differed greatly among the plant functional traits (Table 2, Tables C1 and C2 in Appendix C). WD had the best performance in all three models, with $R^2$ values of 0.64, 0.68 and 0.67 for random forest, boosted regression trees and ensemble model, respectively. SLA and LPC had $R^2$ values greater than 0.45, while LDMC performed the worst, with $R^2$ values below 0.30.
Table 2 Results of plant functional traits for cross-validated $R^2$, NRMSE and MAE for random forest, boosted regression trees and ensemble model.

<table>
<thead>
<tr>
<th>Traits</th>
<th>Random forest</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Boosted regression trees</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Ensemble model</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SLA</td>
<td>0.48</td>
<td>0.22</td>
<td>5.10</td>
<td>0.48</td>
<td>0.20</td>
<td>5.08</td>
<td>0.49</td>
<td>0.21</td>
<td>5.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDMC</td>
<td>0.23</td>
<td>0.21</td>
<td>0.07</td>
<td>0.28</td>
<td>0.18</td>
<td>0.07</td>
<td>0.24</td>
<td>0.20</td>
<td>0.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LNC</td>
<td>0.33</td>
<td>0.19</td>
<td>4.92</td>
<td>0.34</td>
<td>0.18</td>
<td>4.85</td>
<td>0.34</td>
<td>0.19</td>
<td>4.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LPC</td>
<td>0.51</td>
<td>0.24</td>
<td>0.53</td>
<td>0.51</td>
<td>0.22</td>
<td>0.53</td>
<td>0.51</td>
<td>0.27</td>
<td>0.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LA</td>
<td>0.37</td>
<td>0.45</td>
<td>26.76</td>
<td>0.39</td>
<td>0.51</td>
<td>27.47</td>
<td>0.40</td>
<td>0.58</td>
<td>26.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WD</td>
<td>0.64</td>
<td>0.20</td>
<td>0.10</td>
<td>0.68</td>
<td>0.13</td>
<td>0.10</td>
<td>0.67</td>
<td>0.17</td>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SLA, specific leaf area ($m^2$ kg$^{-1}$); LDMC, leaf dry matter content (g g$^{-1}$); LNC, leaf N concentration (mg g$^{-1}$); LPC, leaf P concentration (mg g$^{-1}$); LA, leaf area (cm$^2$); WD, wood density (g cm$^3$); $R^2$, determinate coefficient; NRMSE, normalized root-mean-square error; MAE, mean absolute error.

3.2 Spatial patterns of predicted plant functional traits

There were relatively consistent spatial patterns for SLA, LNC and LPC, with high values in the northeastern and northwestern regions and the southeastern Qinghai-Tibet Plateau, and low values in southwestern China (Figs. 3a, 3c and 3d, Figs. D1, D2, D3, D5 and D6 in Appendix D). SLA and LPC increased with latitude, while LNC did not vary significantly along the latitudinal gradient. For SLA, LNC and LPC, the variability was low among random forest, boosted regression trees and ensemble model, with an overall CV less than 0.3 (Figs. 4a, 4c and 4d).

LDMC values were relatively high in most regions of China, and the low values were mainly located in eastern Yunnan and the Loess Plateau (Fig. 3b, Figs. D1, D2 and D4 in Appendix D). LA showed high values in the northeastern and southern regions (except for the Sichuan Basin), and the southeastern Qinghai-Tibet Plateau (Fig. 3e, Figs. D1, D2 and D7 in Appendix D). The strong latitudinal gradient was observed in LA, where the values decreased with latitude.

The CV values of LPC decreased with latitude, but other traits did not show latitudinal patterns (Fig. 4). The CV values of LA were relatively high, especially in the northwestern region and the Inner Mongolia-Loess Plateau region (Fig. 4e). WD had high values in the northeastern and southern regions (Fig. 2f, Figs. D1, D2 and D8 in Appendix D), while CV values for WD in China were low throughout China (Fig. 4f).
Figure 3. Spatial patterns of predicted plant functional traits in China based on the ensemble model. The grey curves to the right of the maps display trait distribution along with latitude. The white areas represent artificial land cover types. RF, random forest; BRT, boosted regression trees; ensemble, ensemble model; SLA, specific leaf area; LDMC, leaf dry matter content; LNC, leaf N concentration; LPC, leaf P concentration; LA, leaf area; WD, wood density.
Figure 4. The variability in plant functional trait predictions among random forest, boosted regression trees and ensemble model. The grey curves to the right of the maps display coefficient of variation along with latitude. The white areas represent artificial land cover types. SLA, specific leaf area; LDMC, leaf dry matter content; LNC, leaf N concentration; LPC, leaf P concentration; LA, leaf area; WD, wood density.

3.3 Relative importance of predictive variables

The dominant factors explaining spatial variation differed greatly among plant functional traits (Table 3). Overall, climate variables were more important for predicting plant functional traits than were soil variables. Temperature variables (i.e., MAT, MDR and TS) showed close relationships with SLA, LDMC, LPC and WD, while precipitation variables (i.e., PS, PEQ, MAP and PDQ) were more important for predicting the spatial patterns of LNC, LPC and LA. RAD was
the fourth most dominant factor in predicting the spatial patterns of SLA and WD. Elevation also
played an important role in the LDMC and LPC predictions. Within soil variables, soil nutrients
(i.e., pH and SAP) showed close associations with SLA and LNC. In addition to the environmental
variables, MTCI emerged as an important predictor for explaining SLA, LDMC and LA. Finally,
EVI was the most important predictor for LA, and MIR in January and May were the primary
predictors of WD. The relationships between plant functional traits and the most important
variables were shown in Figs. E1 and E2 in Appendix E.

Table 3 List of the eight most important variables for plant functional trait predictions.

<table>
<thead>
<tr>
<th>Rank</th>
<th>SLA</th>
<th>LDMC</th>
<th>LNC</th>
<th>LPC</th>
<th>LA</th>
<th>WD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SAP</td>
<td>MAT</td>
<td>PS</td>
<td>MDR</td>
<td>EVI5</td>
<td>MIR1</td>
</tr>
<tr>
<td>2</td>
<td>TS</td>
<td>Elevation</td>
<td>SAP</td>
<td>PDQ</td>
<td>PEQ</td>
<td>TS</td>
</tr>
<tr>
<td>3</td>
<td>blue9</td>
<td>MTCI5</td>
<td>pH</td>
<td>Elevation</td>
<td>MTCI9</td>
<td>MIR5</td>
</tr>
<tr>
<td>4</td>
<td>RAD</td>
<td>blue8</td>
<td>MDR</td>
<td>MIR8</td>
<td>NIR9</td>
<td>RAD</td>
</tr>
<tr>
<td>5</td>
<td>MTCI4</td>
<td>MTCI4</td>
<td>MAP</td>
<td>Tmax</td>
<td>AI</td>
<td>MIR6</td>
</tr>
<tr>
<td>6</td>
<td>MTCI6</td>
<td>MTCI6</td>
<td>PEQ</td>
<td>MTCI6</td>
<td>MTCI6</td>
<td>pH</td>
</tr>
<tr>
<td>7</td>
<td>Elevation</td>
<td>NIR1</td>
<td>MIR1</td>
<td>MIR7</td>
<td>MAP</td>
<td>red5</td>
</tr>
<tr>
<td>8</td>
<td>MTCI7</td>
<td>CEC</td>
<td>Tmax</td>
<td>MIR9</td>
<td>red5</td>
<td>PS</td>
</tr>
</tbody>
</table>

SLA, specific leaf area (m² kg⁻¹); LDMC, leaf dry matter content (g g⁻¹); LNC, leaf N concentration
(mg g⁻¹); LPC, leaf P concentration (mg g⁻¹); LA, leaf area (cm²); WD, wood density (g cm⁻³); SAP, soil
available P; TS, temperature seasonality; blue, blue reflectance; RAD, solar radiation; MTCI, MERIS
terrestrial chlorophyll index; MAT, mean annual temperature; NIR, near-infrared reflectance; CEC,
cation exchange capacity; PS, precipitation seasonality; MIR, mean diurnal range; MAP, mean annual
precipitation; PEQ, precipitation of wettest quarter of a year; MIR, middle infrared reflectance; Tmax,
max temperature of warmest month of a year; PDQ, precipitation of driest quarter of a year; EVI,
enhanced vegetation index; AI, aridity index; red, red reflectance.

3.4 Model performance

The distributions of the predictive trait values based on random forest, boosted regression trees,
and ensemble model were consistent with the original trait observations, especially the peak
values (Fig. 5). The mean values of trait observations were relatively higher than those of the
predictive values.
Figure 5. Comparison of trait distribution between observations and predictive values in each of the different models. Each panel depicts the distribution of observations in solid red, of the random forest (RF) model in yellow, of the boosted regression trees (BRT) model in blue, and of the ensemble model in green. The dashed vertical lines indicate mean values. SLA, specific leaf area; LDMC, leaf dry matter content; LNC, leaf N concentration; LPC, leaf P concentration; LA, leaf area; WD, wood density.

3.5 Uncertainty assessments

The MESS values of all plant functional traits were positive in most regions, indicating a wide applicability domain of our models (Fig. 6). Nevertheless, trait predictions should be interpreted carefully for northeastern China and the Qinghai-Tibet Plateau due to the sparse samplings in these regions.
Figure 6. Multivariate environmental similarity surface (MESS) assessments for the six plant functional traits. The black dots represented the locations of trait observations. More intense shades indicate greater similarity (blue) or difference (red) in environmental conditions of the location compared to the predictive factors covered by the training dataset. The white areas represent artificial land cover types. SLA, specific leaf area; LDMC, leaf dry matter content; LNC, leaf N concentration; LPC, leaf P concentration; LA, leaf area; WD, wood density.

4 Discussion

4.1 Comparison with previous work

Our study predicted the spatial patterns of six key plant functional traits across China using machine learning methods and identified the applicability domain of the models. WD had the highest precision with an average of $R^2$ of 0.66, which was higher than the global WD prediction (Boonman et al., 2020). This improvement in precision may be attributed to the large number and dense occurrence of sample sites as well as the inclusion of vegetation indices in our study. In addition, SLA and LPC also showed good accuracy with $R^2$ values of 0.50, which was higher than that of Boonman et al. (2020) and consistent with that of Moreno-Martínez et al. (2018). However,
LNC and LA showed relatively poor performance, which may be related to the reason that these two traits were more influenced by phylogeny than environmental variables (Yang et al., 2017; An et al., 2021).

The frequency distribution of plant functional traits in China differed between our study and previous studies (Fig. 7, Fig. F1, Table F1 in Appendix F). Given that the spatial resolution of trait maps in most previous studies is 0.5° (except for Moreno-Martínez et al. (2018) and Vallicrosa et al. (2022)), we resampled the data products of previous studies and our study to 0.5° spatial resolution. The distribution in our study contained more predictions at lower values of SLA, LNC and LPC and was broader than those for SLA and LNC in previous global studies. However, the distribution of LNC in our study was consistent with that in Vallicrosa et al. (2022) at the 1 km spatial resolution (Fig. F1 in Appendix F). LA in our study contained more predictions at higher values and was also broader than those in previous global studies. WD did not show the lower and higher predictive values in this study, however, the WD values in the studies of Boonman et al. (2020) and Schiller et al. (2021) had more predictions at higher values and no lower values (< 0.3 g cm⁻³). Our predicted values of SLA showed the highest spatial correlation with those of Dong et al. (2023), and LNC showed the strongest spatial correlation with those of Butler et al. (2017) (Table 5). LA and WD showed the best spatial correlation with those of Schiller et al. (2021), but LPC showed relatively weak spatial correlation with those of published studies.

In addition, we compared our results to the other studies focused on China. Yang et al. (2016) predicted the spatial distribution of leaf mass per area (i.e., 1/SLA) and LNC based on trait-environment relationships in China and had an R² of 0.13-0.16. The lower predictive precision may be because Yang et al. (2016) only used MAT, MAP and RAD as predictors in estimating the spatial patterns of leaf mass per area and LNC, which likely led to poor performance and low heterogeneity. These results also demonstrated the advantage of our methods in mapping the spatial patterns of plant functional traits at a regional scale.
Table 5  Spatial correlations for SLA, LNC, LPC, LA and WD between this study and other previous trait maps, labelled by the first author of the corresponding publication (see Table F1 in Appendix F for citations)

<table>
<thead>
<tr>
<th>Spatial correlation</th>
<th>Dong</th>
<th>Vallicrosa</th>
<th>Schiller</th>
<th>Boonman</th>
<th>Moreno</th>
<th>Madani</th>
<th>Butler</th>
<th>Bodegom</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLA</td>
<td>0.398</td>
<td>-0.082</td>
<td>0.327</td>
<td>0.242</td>
<td>0.136</td>
<td>-0.042</td>
<td>0.319</td>
<td></td>
</tr>
<tr>
<td>LNC</td>
<td>0.156</td>
<td>0.359</td>
<td>0.229</td>
<td>0.252</td>
<td>0.394</td>
<td>0.057</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LPC</td>
<td></td>
<td>0.136</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LA</td>
<td></td>
<td>0.514</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WD</td>
<td></td>
<td>0.647</td>
<td>0.107</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The spatial correlation of leaf dry matter content (LDMC) between our study and previous studies was not included, as the LDMC maps were not available. SLA, specific leaf area (m² kg⁻¹); LNC, leaf N concentration (mg g⁻¹); LPC, leaf P concentration (mg g⁻¹); LA, leaf area (cm²); WD, wood density (g cm⁻³).

Figure 7. Frequency distributions of plant functional traits in our study (“This study”, dashed black lines) and other trait maps, identified by the first author of the corresponding publication (see Table F1 for citations). SLA, specific leaf area (m² kg⁻¹); LNC, leaf N concentration (mg g⁻¹);
Our study revealed the spatial patterns of different plant functional traits across China, and the variability among the two machine learning methods was relatively low. We compared the spatial differences of trait maps between our study and previous studies at the global scale (Figs. F2-F6 in Appendix F). For example, our study showed high SLA values in the southeastern Qinghai-Tibet Plateau, which concurred with the global study of Boonman et al. (2020). The spatial difference of SLA between our study and Bodegom et al. (2014) was relatively low, and the predictive values in most regions were slightly lower in our study than those in Bodegom et al. (2014). The spatial pattern of difference in SLA between our study and Moreno et al. (2018), Bulter et al. (2017) and Bodegom et al. (2020) was consistent, and the values were higher in northeastern China and southwestern Qinghai-Tibet Plateau in our study than those studies. Our study showed higher LNC values in the northern Inner Mongolia-the Loess Plateau-the eastern Qinghai-Tibet Plateau and northwestern China than those studies at the global studies (Butler et al., 2017; Moreno-Martínez et al., 2018; Boonman et al., 2020; Vallicrosa et al., 2022; Dong et al., 2023), reflecting the consistent spatial pattern among these studies. However, Yang et al. (2016) predicted high LNC values in northeastern and northwestern China, northern Inner Mongolia and the entire Qinghai-Tibet Plateau, and SLA and LNC had low heterogeneity overall. The discrepancy with Yang et al. (2016) may be attributed to spatial extrapolation based on trait-climate relationships with a low predictive precision. There was no consistent spatial pattern in LPC between our study and previous studies. Consistent with the global pattern (Wright et al., 2017), LA was larger in southern regions than in northern regions and showed a decreasing trend with latitude. In addition, LA and WD values in our study were lower in most regions than those ones at the global scale. These discrepancies between our study and previous studies at the global scale may be related to three reasons. First, there is bias in the available in-situ field measurement data from China in these global studies, with large gaps in western China for SLA and no data in China for WD (Boonman et al., 2020). Second, some trait-environment relationships may be scale-dependent (Brueelheide et al., 2018), and these studies we compared are from the global scale because the trait maps in China are not available. Third, the methods used for trait mapping were different among studies, including eco-evolutionary optimality models (Dong et al., 2023), Convolutional Neural Networks based on RGB photographs (Schiller et al., 2021), machine learning algorithms (Vallicrosa et al., 2022; Boonman et al., 2020) and multiple regression analysis (Bodegom et al., 2014).

Moreover, our study also identified the applicability domain of our models for predicting the spatial patterns of plant functional traits across China. Five leaf traits and WD appeared to have poor applicability in northeastern China and the Qinghai-Tibet Plateau, primarily due to sparse samplings. Future studies predicting plant functional traits across a large scale through remote
sensing observations or other supplementary data will be needed to re-evaluate our results.

4.3 The role of predictive variables

Our study indicates that environmental variables are important for predicting the spatial patterns of plant functional traits, especially climate variables. Temperature variables were primary predictors for SLA, LDMC, LPC and WD. The relationships between leaf traits and temperature have been widely discussed in global and regional studies (Reich and Oleksyn, 2004; Bruelheide et al., 2018). The positive linkage between WD and temperature may be driven by changes in water viscosity. Plants can adapt to the low water viscosity at high temperatures by reducing the diameter and density of their vessels and by thickening cell walls (Roderick and Berry, 2002; Thomas et al., 2004). Precipitation variables were important predictors for leaf nutrient traits and LA. For example, precipitation of wettest quarter of a year was the factor that most influenced LA variation, which has been confirmed by a previous study (An et al., 2021). A smaller LA could be an adaptive strategy to decrease water loss via reducing the surface area for transpiration under dry environmental conditions (Du et al., 2019). Although the effects of soil on trait predictions were relatively weak, we found that SAP and pH played key roles in SLA and LNC predictions. These results were similar with the previous studies that reported that soil pH was an important driver of trait variation at the global scale and in tundra regions (Maire et al., 2015; Kemppinen et al., 2021). Additionally, from the perspective of cost-efficient theory, the strong effects of SAP reflected that high SLA may be an adaptation for facilitating soil exploration more efficiently in fertile soils (Freschet et al., 2010).

Vegetation indices have recently been proposed as important predictors of spatial patterns of plant functional traits (Loozen et al., 2018). Our results corroborated these findings and further suggested that EVI, MTCI and MIR reflectance are important predictors in models. Here, the underlying mechanisms between vegetation indices and plant functional traits are not further discussed due to their complexity and uncertainty. However, our results indicated that vegetation indices and NIR reflectance are not key predictors of LNC estimation, which contrasts the findings from global and regional studies (Wang et al., 2016; Loozen et al., 2018; Moreno-Martínez et al., 2018). This may be related to the multitude of factors that influence the relationship between LNC and vegetation indices and NIR reflectance, such as forest type and canopy structure (Dahlin et al., 2013).

4.4 Uncertainties

Although our study mapped the spatial patterns of key functional traits of seed plants in China through large-scale field investigations and compared the predictions with previous studies performed at global and regional scales, there persists some uncertainties in the interpretation of these results. First, the predictive ability of models was relatively worse for certain traits, especially LDMC. Beyond the environmental effects, the variation in plant functional traits is also
regulated by phylogenetic structure among plant species (e.g., family, order and phylogenetic
clade) (Li et al., 2017). Consequently, incorporating the phylogenetic information will be a
promising avenue for further improving the accuracy of spatial predictions of plant functional
traits (Butler et al., 2017). A second potential issue is sampling bias; there were major spatial gaps
in field investigation in both the northeastern China and the Qinghai-Tibet Plateau. Due to the few
measurements for shrubs and the lack of herbs, WD data is mainly confined to eastern forests, and
the overall quantity of WD data was much lower than that of leaf traits, even in the TRY database.
The environmental information of sampling sites was not always obtained from original literature,
thus using the public environmental products is a common resolution in large-scale plant trait
studies (Boonman et al., 2020; Vallicrosa et al., 2022). Such mismatch between in-situ trait
measurements and predictors should be resolved in further work. Finally, additional key
challenges in data availability must be resolved to scale up from the species to the community
levels, in particular with data surrounding species co-occurrence and their relative cover or
abundance in ecological communities (He et al., 2023). For example, Global biodiversity data
(e.g., sPlot and Global Biodiversity Information Agency databases) that contains information on
species occurrence or the proportion of species in a community has the potential for enabling the
calculation of community-weighted trait values and the re-evaluation of our results in future work
(Telenius, 2011; Bruehlheide et al., 2019). The lack of consistent time period and spatial resolution
of predictors due to limitation of data availability is another key challenge in the spatial mapping
of plant functional traits. In addition, although WorldClim version 2.1 product has high spatial
resolution and includes various aspects of climatic parameters, there exists certain limitation and
uncertainty in predicting trait maps. Therefore, integrating satellite remote sensing monitoring
methods with in-situ trait data collection can also provide an effective way to estimate and assess
the species diversity at large scales (Cavender-Bares et al., 2022).

4.5 Potential applications

Maps of these key functional traits of seed plants highlighted large-scale variability in space,
which will significantly advance ecological analyses and future interdisciplinary research. First,
using the spatially continuous trait maps, one can optimize and develop trait-flexible vegetation
models to reduce uncertainties of conventional vegetation models based on PFTs, which allows for
the exploration of the community assembly rules based on how plants with different trait
combinations perform under a given set of environmental conditions (Berzaghi et al., 2020). When
trait-flexible vegetation models are available, incorporating trait maps into models will bridge the
gap for vegetation classifications and predictions of vegetation distribution under global change
(Van Bodegom et al., 2012; Yang et al., 2019). Second, most studies focused on the effects of plant
functional traits on ecosystem carbon processes at individual, species and community scales, while
how such effects scale up to regional or larger scales remains challenging. In addition, the
assessments of China’s terrestrial ecosystem carbon sink have had large uncertainties so far (Piao
et al., 2022). The spatial continuous trait maps will provide an effective way to link ecosystem characteristics to ecosystem carbon sink estimates in China (Madani et al., 2018; Šímová et al., 2019). These analyses will help shed light on the mechanisms underlying plant functional traits and terrestrial ecosystem carbon storage at a large scale.

5 Data availability

The original plant functional trait data collected in this study that were used for machine learning models (named by Data file used for machine learning models.csv) and final maps of plant functional traits in terrestrial ecosystems in a GeoTIFF format across China (named by plant functional trait category) are now available for the private link https://figshare.com/s/c527c12d310cb8156ed2 (An et al., 2023). Once the article is accepted, we will publicly publish these maps at the figshare website.

6 Conclusions

We generated a set of spatial continuous trait maps at a 1-km spatial resolution using machine learning methods in combination with field measurements, environmental variables and vegetation indices. Models for leaf traits (except for LDMC) and WD showed good accuracy and robustness, whereas models of LDMC had relatively poor precision and robustness. Temperature variables were the most important predictors for leaf traits (except for LA) and WD, and precipitation variables were the most important predictors for leaf nutrient traits and LA. We caution that plant functional trait predictions should be interpreted carefully for northeastern China and the Qinghai-Tibet Plateau. The spatial continuous trait maps generated in our study are complementary to current terrestrial in-situ observations and offer new avenues for predicting large-scale changes in vegetation and ecosystem function under climate scenarios in China.

Appendix A Data collection from literature


Feng Q. H. The study on the response of foliar δ13C of different life from plants to altitude in subalpine area of Western Sichuan, China. 2011, Graduation Thesis.


Guo W. J. Exploring the relationship between arbuscular mycorrhizal fungi and plant based on phylogeny and plant traits. 2015, Graduation Thesis.


He Y. T. Studies on physioecological traits of 30 plant species in the Subalpine Meadow of the Qinling Mountains. 2007, Graduation Thesis.

Hou M M. Adaptive evolution of some species from sedges (Carex Cyperaceae) based on phylogeny and leaf functional traits to habitat in the Poyang Lake Area. 2017, Graduation Thesis.


Huang Y. L. The research about the turnover patterns and moisture adaptation mechanism of major species on the South-North-facing slope. 2012, Graduation Thesis.


Jia Q. Q. Functional traits of fine roots and their relationship with leaf traits of 50 major species in a subtropical forest in Gutianshan. 2011, Graduation Thesis.


Jing G. H. Responses of grassland community structure and functions to management practices on the semi-arid area of Loess Plateau. 2017, Graduation Thesis.


Li F. Comparison of functional traits in semi-humid evergreen broad-leaved in Western Hill of Kunming. 2011, Graduation Thesis.


Liu C. H. and Li Y. Y. Relationship between leaf traits and PV curve parameters in the typical deciduous woody plants occurring in Southern Huanglong Mountain. Journal of Northwest Forestry University, 2013, 28, 1-5.


Qi L. X. Response of leaf traits of *Pinus mongoliensis* and *Pinus massoniana* to elevation gradient in Daiyun Mountain. 2015, Graduation Thesis.


Shang K. K. Differentiation and maintenance of relict deciduous broad-leaved forest patterns along microtopographic gradient in subtropical area, East China. 2011, Graduation Thesis.


Tan X. Y. Research on leaf functional diversity of forest communities in rainy area of south-west China. 2014,
Graduation Thesis.


Wang S. S. The traits and adaptive strategies of main herbaceous plants and lianas on micro-topographical units in Longishan reserves of Anhui Province. 2016, Graduation Thesis.


Xu Y. Z. Biomass estimate and storage mechanisms in northern subtropical forest ecosystems, central China. 2016,
Graduation Thesis.


Yang S. The adaptive strategies of main herbaceous plants traits to different micro-topographical units in Dashushan Mountain, Hefei. 2017, Graduation Thesis.


Yuan S. Preliminary research on plant functional traits and the capability of carbon sequestration of major tree species in Changbai Mountain Area. 2011, Graduation Thesis.


Zhang J. Y. Comparative study on the different plant functional groups leaf traits at the Maoershan Region. 2008, Graduation Thesis.


## Appendix B

### Table B1 Summary of statistics in plant functional traits, environmental variables and geographical distribution in China.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Unit</th>
<th>Range</th>
<th>Mean</th>
<th>CV (%)</th>
<th>No. of species</th>
<th>Entries</th>
<th>Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLA</td>
<td>m² kg⁻¹</td>
<td>0.06–81.68</td>
<td>17.88</td>
<td>54.96</td>
<td>2463</td>
<td>9195</td>
<td>1032</td>
</tr>
<tr>
<td>LDMC</td>
<td>g g⁻¹</td>
<td>0.06–0.95</td>
<td>0.34</td>
<td>100.00</td>
<td>1582</td>
<td>3957</td>
<td>193</td>
</tr>
<tr>
<td>LNC</td>
<td>mg g⁻¹</td>
<td>3.41–66.02</td>
<td>21.52</td>
<td>37.44</td>
<td>2335</td>
<td>7407</td>
<td>567</td>
</tr>
<tr>
<td>LPC</td>
<td>mg g⁻¹</td>
<td>0.09–9.70</td>
<td>1.83</td>
<td>62.19</td>
<td>2074</td>
<td>6266</td>
<td>515</td>
</tr>
<tr>
<td>LA</td>
<td>cm²</td>
<td>0.0033–2553.33</td>
<td>36.16</td>
<td>259.64</td>
<td>1838</td>
<td>5976</td>
<td>691</td>
</tr>
<tr>
<td>WD</td>
<td>g cm⁻³</td>
<td>0.25–1.37</td>
<td>0.68</td>
<td>33.16</td>
<td>768</td>
<td>1788</td>
<td>639</td>
</tr>
<tr>
<td>Altitude</td>
<td>m</td>
<td>-144–5454</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAT</td>
<td>°C</td>
<td>-12.07–24.32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAP</td>
<td>mm</td>
<td>15–2982</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soil total N</td>
<td>g kg⁻¹</td>
<td>0.11–10.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bulk density</td>
<td>g cm⁻³</td>
<td>0.83–1.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SLA, specific leaf area; LDMC, leaf dry matter content; LNC, leaf N concentration; LPC, leaf P concentration; LA, leaf area; WD, wood density; MAT, mean annual temperature; MAP, mean annual precipitation.
Table B2 List of all the predictors including environment and remote sensing variables used in this study.

<table>
<thead>
<tr>
<th>Type of variables</th>
<th>Variable name</th>
<th>Abbreviations</th>
<th>Units</th>
<th>Time periods</th>
<th>Spatial resolution</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Climate</td>
<td>Mean annual temperature</td>
<td>MAT</td>
<td>°C</td>
<td>1970-2000</td>
<td>1 km</td>
<td>WorldClim version 2.1</td>
</tr>
<tr>
<td></td>
<td>Mean diurnal range</td>
<td>MDR</td>
<td>°C</td>
<td>1970-2000</td>
<td>1 km</td>
<td>WorldClim version 2.1</td>
</tr>
<tr>
<td></td>
<td>Temperature seasonality</td>
<td>TS</td>
<td>°C</td>
<td>1970-2000</td>
<td>1 km</td>
<td>WorldClim version 2.1</td>
</tr>
<tr>
<td></td>
<td>Max temperature of warmest month</td>
<td>Tmax</td>
<td>°C</td>
<td>1970-2000</td>
<td>1 km</td>
<td>WorldClim version 2.1</td>
</tr>
<tr>
<td></td>
<td>Min temperature of coldest month</td>
<td>Tmin</td>
<td>°C</td>
<td>1970-2000</td>
<td>1 km</td>
<td>WorldClim version 2.1</td>
</tr>
<tr>
<td></td>
<td>Temperature annual range</td>
<td>TAR</td>
<td>°C</td>
<td>1970-2000</td>
<td>1 km</td>
<td>WorldClim version 2.1</td>
</tr>
<tr>
<td></td>
<td>Isothermality</td>
<td>IS</td>
<td>%</td>
<td>1970-2000</td>
<td>1 km</td>
<td>WorldClim version 2.1</td>
</tr>
<tr>
<td></td>
<td>Mean temperature of wettest quarter</td>
<td>MTEQ</td>
<td>°C</td>
<td>1970-2000</td>
<td>1 km</td>
<td>WorldClim version 2.1</td>
</tr>
<tr>
<td></td>
<td>Mean temperature of driest quarter</td>
<td>MTDQ</td>
<td>°C</td>
<td>1970-2000</td>
<td>1 km</td>
<td>WorldClim version 2.1</td>
</tr>
<tr>
<td></td>
<td>Mean temperature of warmest quarter</td>
<td>MTWQ</td>
<td>°C</td>
<td>1970-2000</td>
<td>1 km</td>
<td>WorldClim version 2.1</td>
</tr>
<tr>
<td></td>
<td>Mean temperature of coldest quarter</td>
<td>MTCQ</td>
<td>°C</td>
<td>1970-2000</td>
<td>1 km</td>
<td>WorldClim version 2.1</td>
</tr>
<tr>
<td></td>
<td>Mean annual precipitation</td>
<td>MAP</td>
<td>mm</td>
<td>1970-2000</td>
<td>1 km</td>
<td>WorldClim version 2.1</td>
</tr>
<tr>
<td></td>
<td>Precipitation of wettest month</td>
<td>PEM</td>
<td>mm</td>
<td>1970-2000</td>
<td>1 km</td>
<td>WorldClim version 2.1</td>
</tr>
<tr>
<td></td>
<td>Precipitation of driest month</td>
<td>PDM</td>
<td>mm</td>
<td>1970-2000</td>
<td>1 km</td>
<td>WorldClim version 2.1</td>
</tr>
<tr>
<td></td>
<td>Precipitation seasonality</td>
<td>PS</td>
<td>%</td>
<td>1970-2000</td>
<td>1 km</td>
<td>WorldClim version 2.1</td>
</tr>
<tr>
<td></td>
<td>Precipitation of wettest quarter</td>
<td>PEQ</td>
<td>mm</td>
<td>1970-2000</td>
<td>1 km</td>
<td>WorldClim version 2.1</td>
</tr>
<tr>
<td></td>
<td>Precipitation of driest quarter</td>
<td>PDQ</td>
<td>mm</td>
<td>1970-2000</td>
<td>1 km</td>
<td>WorldClim version 2.1</td>
</tr>
<tr>
<td></td>
<td>Precipitation of warmest quarter</td>
<td>PWQ</td>
<td>mm</td>
<td>1970-2000</td>
<td>1 km</td>
<td>WorldClim version 2.1</td>
</tr>
<tr>
<td></td>
<td>Precipitation of coldest quarter</td>
<td>PCQ</td>
<td>mm</td>
<td>1970-2000</td>
<td>1 km</td>
<td>WorldClim version 2.1</td>
</tr>
<tr>
<td></td>
<td>Aridity index</td>
<td>AI</td>
<td>/</td>
<td>1970-2000</td>
<td>1 km</td>
<td>Global CGIAR-CSI</td>
</tr>
<tr>
<td></td>
<td>Solar radiation</td>
<td>RAD</td>
<td>kJ m⁻² day⁻¹</td>
<td>1970-2000</td>
<td>1 km</td>
<td>WorldClim version 2.1</td>
</tr>
<tr>
<td>Topography</td>
<td>Elevation</td>
<td>/</td>
<td>m</td>
<td>1970-2000</td>
<td>1 km</td>
<td>SRTM 90m V4.1</td>
</tr>
<tr>
<td>Soil</td>
<td>Soil sand content</td>
<td>SAND</td>
<td>%</td>
<td>/</td>
<td>1 km</td>
<td>Shangguan et al. (2013)</td>
</tr>
<tr>
<td></td>
<td>Soil silt content</td>
<td>SILT</td>
<td>%</td>
<td>/</td>
<td>1 km</td>
<td>Shangguan et al. (2013)</td>
</tr>
<tr>
<td></td>
<td>Soil clay content</td>
<td>CLAY</td>
<td>%</td>
<td>/</td>
<td>1 km</td>
<td>Shangguan et al. (2013)</td>
</tr>
<tr>
<td></td>
<td>Bulk density</td>
<td>BD</td>
<td>g cm⁻³</td>
<td>/</td>
<td>1 km</td>
<td>Shangguan et al. (2013)</td>
</tr>
<tr>
<td></td>
<td>Soil pH</td>
<td>pH</td>
<td>/</td>
<td>/</td>
<td>1 km</td>
<td>Shangguan et al. (2013)</td>
</tr>
<tr>
<td></td>
<td>Soil organic matter</td>
<td>SOC</td>
<td>g kg⁻¹</td>
<td>/</td>
<td>1 km</td>
<td>Shangguan et al. (2013)</td>
</tr>
<tr>
<td></td>
<td>Soil total N</td>
<td>STN</td>
<td>g kg⁻¹</td>
<td>/</td>
<td>1 km</td>
<td>Shangguan et al. (2013)</td>
</tr>
<tr>
<td></td>
<td>Soil total P</td>
<td>STP</td>
<td>g kg⁻¹</td>
<td>/</td>
<td>1 km</td>
<td>Shangguan et al. (2013)</td>
</tr>
<tr>
<td></td>
<td>Soil alkali-hydrolysable N</td>
<td>SAN</td>
<td>mg kg⁻¹</td>
<td>/</td>
<td>1 km</td>
<td>Shangguan et al. (2013)</td>
</tr>
<tr>
<td></td>
<td>Soil available P</td>
<td>SAP</td>
<td>mg kg⁻¹</td>
<td>/</td>
<td>1 km</td>
<td>Shangguan et al. (2013)</td>
</tr>
<tr>
<td></td>
<td>Soil available K</td>
<td>SAK</td>
<td>mg kg⁻¹</td>
<td>/</td>
<td>1 km</td>
<td>Shangguan et al. (2013)</td>
</tr>
<tr>
<td></td>
<td>Cation exchange capacity</td>
<td>CEC</td>
<td>me kg⁻¹</td>
<td>/</td>
<td>1 km</td>
<td>Shangguan et al. (2013)</td>
</tr>
</tbody>
</table>
The remote sensing variables are calculated as long-term monthly averages from 2001 to 2018. Thus 12 variables of each remote sensing category are obtained.

Table B3 The number of samples of eight plant functional trait used for model training (80%) and validation (20%).

<table>
<thead>
<tr>
<th>Traits</th>
<th>No. of samples</th>
<th>No. of samples used for model training</th>
<th>No. of samples used for model validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLA</td>
<td>9195</td>
<td>7356</td>
<td>1839</td>
</tr>
<tr>
<td>LDMC</td>
<td>3957</td>
<td>3166</td>
<td>791</td>
</tr>
<tr>
<td>LNC</td>
<td>7407</td>
<td>5926</td>
<td>1481</td>
</tr>
<tr>
<td>LPC</td>
<td>6266</td>
<td>5013</td>
<td>1253</td>
</tr>
<tr>
<td>LA</td>
<td>5976</td>
<td>4781</td>
<td>1195</td>
</tr>
<tr>
<td>WD</td>
<td>1787</td>
<td>1430</td>
<td>357</td>
</tr>
</tbody>
</table>

SLA, specific leaf area (m² kg⁻¹); LDMC, leaf dry matter content (g g⁻¹); LNC, leaf N concentration (mg g⁻¹); LPC, leaf P concentration (mg g⁻¹); LA, leaf area (cm²); WD, wood density (g cm⁻³).
Figure B1. Correlations among climate variables. The blank indicates that the correlations are not significant ($P > 0.05$). The size of the circles is proportional to the correlation coefficient. The abbreviation of climate variables is seen in Table B2.
Figure B2. Correlations among soil variables. The blank indicates that the correlations are not significant ($P > 0.05$). The size of the circles is proportional to the correlation coefficient. The abbreviation of soil variables is seen in Table B2.
Figure B3. Correlations among monthly remote sensing variables. The blank indicates that the correlations are not significant ($P > 0.05$). The size of the circles is proportional to the correlation coefficient. (a) enhanced vegetation index (EVI); (b) MERIS terrestrial chlorophyll index (MTCI); (c) MIR reflectance; (d) NIR reflectance; (e) red reflectance; (f) blue reflectance.
## Appendix C

### Table C1 Optimal parameter combination and model performance of random forest for plant functional traits

<table>
<thead>
<tr>
<th>Traits</th>
<th>ntree</th>
<th>mtry</th>
<th>$R^2$</th>
<th>NRMSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLA</td>
<td>1000</td>
<td>24</td>
<td>0.476</td>
<td>0.22</td>
<td>5.134</td>
</tr>
<tr>
<td>LDMC</td>
<td>1000</td>
<td>11</td>
<td>0.234</td>
<td>0.20</td>
<td>0.072</td>
</tr>
<tr>
<td>LNC</td>
<td>1000</td>
<td>57</td>
<td>0.392</td>
<td>0.00</td>
<td>0.098</td>
</tr>
<tr>
<td>LPC</td>
<td>1000</td>
<td>20</td>
<td>0.587</td>
<td>0.05</td>
<td>0.129</td>
</tr>
<tr>
<td>LA</td>
<td>1000</td>
<td>18</td>
<td>0.278</td>
<td>0.48</td>
<td>26.622</td>
</tr>
<tr>
<td>WD</td>
<td>1000</td>
<td>9</td>
<td>0.531</td>
<td>0.02</td>
<td>0.072</td>
</tr>
</tbody>
</table>

SLA, specific leaf area; LDMC, leaf dry matter content; LNC, leaf N concentration; LPC, leaf P concentration; LA, leaf area; WD, wood density.

### Table C2 Optimal parameter combination and model performance of boosted regression trees for plant functional traits

<table>
<thead>
<tr>
<th>Traits</th>
<th>n.tree</th>
<th>interaction. depth</th>
<th>shrinkage</th>
<th>learning rate</th>
<th>bag fractions</th>
<th>$R^2$</th>
<th>NRMSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLA</td>
<td>3000</td>
<td>6</td>
<td>0.01</td>
<td>10</td>
<td>0.75</td>
<td>0.486</td>
<td>0.20</td>
<td>5.082</td>
</tr>
<tr>
<td>LDMC</td>
<td>3000</td>
<td>2</td>
<td>0.01</td>
<td>10</td>
<td>0.75</td>
<td>0.247</td>
<td>0.19</td>
<td>0.071</td>
</tr>
<tr>
<td>LNC</td>
<td>3000</td>
<td>6</td>
<td>0.01</td>
<td>10</td>
<td>0.70</td>
<td>0.414</td>
<td>0.00</td>
<td>0.096</td>
</tr>
<tr>
<td>LPC</td>
<td>3000</td>
<td>7</td>
<td>0.01</td>
<td>10</td>
<td>0.75</td>
<td>0.591</td>
<td>0.05</td>
<td>0.129</td>
</tr>
<tr>
<td>LA</td>
<td>3000</td>
<td>3</td>
<td>0.001</td>
<td>10</td>
<td>0.75</td>
<td>0.282</td>
<td>0.55</td>
<td>27.556</td>
</tr>
<tr>
<td>WD</td>
<td>3000</td>
<td>4</td>
<td>0.01</td>
<td>10</td>
<td>0.70</td>
<td>0.627</td>
<td>0.01</td>
<td>0.066</td>
</tr>
</tbody>
</table>

SLA, specific leaf area; LDMC, leaf dry matter content; LNC, leaf N concentration; LPC, leaf P concentration; LA, leaf area; WD, wood density.
Figure D1. Spatial distribution of plant functional traits based on random forest. The grey curves on the right of maps were trait distribution along with latitude. The white areas represent artificial land cover types. SLA, specific leaf area; LDMC, leaf dry matter content; LNC, leaf N concentration; LPC, leaf P concentration; LA, leaf area; WD, wood density.
**Figure D2.** Spatial distribution of plant functional traits based on boosted regression trees. The grey curves on the right of maps were trait distribution along with latitude. The white areas represent artificial land cover types. SLA, specific leaf area; LDMC, leaf dry matter content; LNC, leaf N concentration; LPC, leaf P concentration; LA, leaf area; WD, wood density.
Figure D3. Spatial distribution of specific leaf area for each plant functional type. The left pen was obtained from RF method (random forest), the right pen was obtained from BRT method (boosted regression trees). The white areas represent other natural vegetation types and artificial land cover types. EBF, evergreen broadleaf forest; ENF, evergreen needleleaf forest; DBF, deciduous broadleaf forest; DNF, deciduous needleleaf forest; SHRUB, shrubland; GRASS, grassland.
Figure D4. Spatial distribution of leaf dry matter content for each plant functional type. The left panel was obtained from RF method (random forest), the right panel was obtained from BRT method (boosted regression trees). The white areas represent other natural vegetation types and artificial land cover types. EBF, evergreen broadleaf forest; ENF, evergreen needleleaf forest; DBF, deciduous broadleaf forest; DNF, deciduous needleleaf forest; SHRUB, shrubland; GRASS, grassland.
**Figure D5.** Spatial distribution of leaf N concentration for each plant functional type. The left penal was obtained from RF method (random forest), the right penal was obtained from BRT method (boosted regression trees). The white areas represent other natural vegetation types and artificial land cover types. EBF, evergreen broadleaf forest; ENF, evergreen needleleaf forest; DBF, deciduous broadleaf forest; DNF, deciduous needleleaf forest; SHRUB, shrubland; GRASS, grassland.
Figure D6. Spatial distribution of leaf P concentration for each plant functional type. The left penal was obtained from RF method (random forest), the right penal was obtained from BRT method (boosted regression trees). The white areas represent other natural vegetation types and artificial land cover types. EBF, evergreen broadleaf forest; ENF, evergreen needleleaf forest; DBF, deciduous broadleaf forest; DNF, deciduous needleleaf forest; SHRUB, shrubland; GRASS, grassland.
Figure D7. Spatial distribution of leaf area for each plant functional type. The left penal was obtained from RF method (random forest), the right penal was obtained from BRT method (boosted regression trees). The white areas represent other natural vegetation types and artificial land cover types. EBF, evergreen broadleaf forest; ENF, evergreen needleleaf forest; DBF, deciduous broadleaf forest; DNF, deciduous needleleaf forest; SHRUB, shrubland; GRASS, grassland.
Figure D8. Spatial distribution of wood density for each plant functional type. The left penal was obtained from RF method (random forest), the right penal was obtained from BRT method (boosted regression trees). The white areas represent other natural vegetation types and artificial land cover types. EBF, evergreen broadleaf forest; ENF, evergreen needleleaf forest; DBF, deciduous broadleaf forest; DNF, deciduous needleleaf forest; SHRUB, shrubland.
Figure E1. The relationships between SLA (specific leaf area), LDMC (leaf dry matter content), LNC (leaf N concentration), LPC (leaf P concentration) and their eight most important predictors.
Figure E2. The relationships between LA (leaf area), WD (wood density) and their eight most important predictors.
Appendix F Comparisons between our study with trait maps from previous studies

Given that the trait maps predicted for China were not available from the literature and authors, we compared our study with those studies performed at the global scale (see Table F1). Thus, we extracted the data in China from global trait maps. Before the quantitative comparisons with previous studies, we performed two steps to make the data products as comparable as possible and improve the consistency between different studies. First, due to different spatial resolution of global trait maps (mainly 0.5°) and our study, we resampled the data products of previous studies and our maps to 0.5° spatial resolution. In addition, Vallicrosa et al. (2022) generated the global maps of LNC and LPC with a 1 km spatial resolution, we also compared the frequency distribution of Vallicrosa et al. (2022) with that of our study at a 1 km spatial resolution. Second, our study focused on natural vegetation, so the global trait maps were used to filter out non-natural vegetation (e.g., croplands). For example, Madani et al. (2018) predicted the spatial distributions of SLA that included croplands. We quantitatively compared our maps with previous studies from two perspectives. The comparisons among trait maps were made using frequency plots and spatial correlations (Figure 7 and Table 5). And the maps of spatial differences between our study and previous studies were displayed as Figs F1-F5 in Appendix F.

Table F1 Summary table of related trait maps of previous studies used in this study.

<table>
<thead>
<tr>
<th>References</th>
<th>Related traits</th>
<th>Methods</th>
<th>Predictors</th>
<th>Consideration of PFT</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dong et al. (2023)</td>
<td>SLA</td>
<td>Optimality models</td>
<td>Climate</td>
<td>Yes</td>
<td>0.5°</td>
</tr>
<tr>
<td></td>
<td>LNC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vallicrosa et al.</td>
<td>LNC</td>
<td>Neural networks</td>
<td>Climate</td>
<td>Yes</td>
<td>0.0083°</td>
</tr>
<tr>
<td>(2022)</td>
<td>LPC</td>
<td></td>
<td>Soil N and P deposition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schiller et al.</td>
<td>SLA</td>
<td>Convolutional Neural Networks</td>
<td>Climate</td>
<td>No</td>
<td>0.5°</td>
</tr>
<tr>
<td>(2021)</td>
<td>LNC</td>
<td></td>
<td>In-situ RGB images</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>WD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boonman et al.</td>
<td>SLA</td>
<td>Generalized linear model, Generalized additive model, Random forest, Boosted regression trees, Ensemble model</td>
<td>Climate</td>
<td>No</td>
<td>0.5°</td>
</tr>
<tr>
<td>(2020)</td>
<td>LNC</td>
<td></td>
<td>Soil</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>WD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moreno et al.</td>
<td>SLA</td>
<td>Regularized linear regression, Random forest, Neural networks, Kernel networks</td>
<td>Climate</td>
<td>Yes</td>
<td>0.0045°</td>
</tr>
<tr>
<td>(2018)</td>
<td>LNC</td>
<td></td>
<td>Elevation Reflectance</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LPC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LDMC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reference</td>
<td>Trait</td>
<td>Model Type</td>
<td>Type</td>
<td>Include</td>
<td>Resolution</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------</td>
<td>---------------------</td>
<td>-------</td>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>Madani et al.</td>
<td>SLA</td>
<td>Generalized additive</td>
<td>Climate</td>
<td>No</td>
<td>0.5°</td>
</tr>
<tr>
<td>(2018)</td>
<td></td>
<td>model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Butler et al.</td>
<td>SLA</td>
<td>Bayesian model</td>
<td>Climate</td>
<td>Yes</td>
<td>0.5°</td>
</tr>
<tr>
<td>(2017)</td>
<td></td>
<td>LNC, LPC</td>
<td>Soil</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bodegom et al.</td>
<td>SLA</td>
<td>Multiple regression</td>
<td>Climate</td>
<td>No</td>
<td>0.5°</td>
</tr>
<tr>
<td>(2014)</td>
<td></td>
<td>WD analysis</td>
<td>Soil</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The resolutions 0.5°, 0.0083° and 0.0045° correspond to square grid cell sizes of about 50 km, 1 km and 500 m at the equator. PFT, plant functional type; SLA, specific leaf area; LDMC, leaf dry matter content; LNC, leaf N concentration; LPC, leaf P concentration; LA, leaf area; WD, wood density.

**Figure F1.** Frequency distributions of plant functional traits in our study (“This study”, dashed black lines) and Vallicrosa et al. (2022) at 1 km spatial resolution. (a) LNC, leaf N concentration (mg g⁻¹); (b) LPC, leaf P concentration (mg g⁻¹).
Figure F2. Spatial differences in SLA (specific leaf area, $\text{m}^2 \text{kg}^{-1}$) between our study and trait maps from previous studies (see Table F1 for citations).
Figure F3. Spatial differences in LNC (leaf N concentration, mg g⁻¹) between our study and trait maps from previous studies (see Table F1 for citations).
Figure F4. Spatial differences in LPC (leaf P concentration, mg g⁻¹) between our study and trait maps from previous studies (see Table F1 for citations).
Figure F5. Spatial differences in LA (leaf area, cm²) between our study and trait maps from previous studies (see Table F1 for citations).
Figure F6. Spatial differences in WD (wood density, g cm$^{-3}$) between our study and trait maps from previous studies (see Table F1 for citations).
Author contributions. NA and NL designed the research. NA did the analysis, processed the data and wrote the draft of the paper. All co-authors commented on the manuscript and agreed upon the final version of the paper.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher’s note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgement. We acknowledge financial supports from the National Natural Science Foundation of China (41991234) and the Joint CAS-MPG Research Project (HZXM20225001MI).

Financial support. This work has been supported by the National Natural Science Foundation of China (grant no. 41991234) and the Joint CAS-MPG Research Project (grant no. HZXM20225001MI).

References


Bohner, A. Soil chemical properties as indicators of plant species richness in grassland communities. Integrating efficient grassland farming and biodiversity, Proceedings of the 13th International Occasional Symposium of the European Grassland Federation, Tartu,


Loozen, Y., Rebel, K. T., de Jong, S. M., Lu, M., Ollinger, S. V., Wassen, M. J., and Karssenberg,


H., Zhang, Q., Chen, D. D., Chen, M., Chu, J. T., Dou, Y. J., Guo, J. X., Li, H. Q., Li, J. J.,
soil properties for land surface modeling. J. Adv. Model. Earth Syst., 5, 212–224,

Siefert, A., Violle, C., Chalmandrier, L., Albert, C. H., Taudiere, A., Fajardo, A., Aarssen, L. W.,
Baraloto, C., Carlucci, M. B., Cianciaruso, M. V., de, L. D. V., de Bello, F., Duarte, L. D.,
Fonseca, C. R., Freschet, G. T., Gaucherand, S., Gross, N., Hikosaka, K., Jackson, B., Jung,
V., Katiyama, C., Katabuchi, M., Kemel, S. W., Kichenin, E., Kraft, N. J., Lagerstrom, A.,
Bagousse-Pinguet, Y. L., Li, Y., Mason, N., Messier, J., Nakashizuka, T., Overton, J. M.,
Peltzer, D. A., Perez-Ramos, I. M., Pillar, V. D., Prentice, H. C., Richardson, S., Sasaki, T.,
Schamp, B. S., Schob, C., Shipley, B., Sundqvist, M., Sykes, M. T., Vandewalle, M., and
Wardle, D. A.: A global meta-analysis of the relative extent of intraspecific trait variation in

Šímová, I., Sandel, B., Enquist, B. J., Michaletz, S. T., Kattge, J., Violle, C., McGill, B. J., Blonder,
B., Engemann, K., Peet, R. K., Wiser, S. K., Morueta-Holme, N., Boyle, B., Kraft, N. J. B.,
Svenning, J. C., and Hector, A.: The relationship of woody plant size and leaf nutrient content
to large-scale productivity for forests across the Americas. J. Ecol., 107, 2278–2290,

Sitch, S., Huntingford, C., Gedney, N., Levy, P. E., Lomas, M., Piao, S. L., Betts, R., Ciais, P., Cox,
terrestrial carbon cycle, future plant geography and climate-carbon cycle feedbacks using five
Dynamic Global Vegetation Models (DGVMs). Global Change Biol., 14, 2015–2039,

Smart, S. M., Glanville, H. C., Blanes, M. d. C., Mercado, L. M., Emmett, B. A., Jones, D. L.,
Cosby, B. J., Marrs, R. H., Butler, A., Marshall, M. R., Reinsch, S., Herrero-Jáuregui, C.,
Hodgson, J. G., and Field, K.: Leaf dry matter content is better at predicting above-ground
net primary production than specific leaf area. Funct. Ecol., 31, 1336–1344,


Thomas, D. S., Montagu, K. D., and Conroy, J. P.: Changes in wood density of Eucalyptus
 camaldulensis due to temperature-the physiological link between water viscosity and wood
2004.

Thomas, S. C.: Photosynthetic capacity peaks at intermediate size in temperate deciduous trees.

Thuiller, W., Lafourcade, B., Engler, R., and Araújo, M. B.: BIOMOD – A platform for ensemble
forecasting of species distributions. Ecography, 32, 369–373, https://doi.org/10.1111/j.1600-


