Reaping the rewards of the big data revolution in ecology

Stuart Ian Graham

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Janneke Hille Ris Lambers, Chair

Joseph Ammirati

Brian Harvey

Program Authorized to Offer Degree:

Biology

University of Washington

Abstract

Reaping the rewards of the big data revolution in ecology

Stuart Ian Graham

Chair of Supervisory Committee:

Janneke Hille Ris Lambers

Department of Biology

The study of ecology has been irreversibly transformed by the ongoing big data revolution. Technological advancements in data collection and a cultural shift toward data sharing mean that ecological data are more available than ever before. However, availability is not equivalent to accessibility; analytical methods in ecology are typically highly customized for specific datasets, and their resulting complexity makes it difficult to adapt them to new analyses. Consequently, ecologists spend a lot of time developing their own analytical tools rather than building on the work of others. To take full advantage of ecology's big data revolution we need tools that are easy to implement and adapt to new use cases. The first part of this dissertation presents a series of such tools designed for the data collected in mapped forest stands; areas of the forest where the precise locations of all trees are recorded. In chapter one, I develop a new regularized regression model of tree growth and compare it to the most popular tree growth model. I find that my model reaches the same conclusions regarding important ecological hypotheses as the classical model, but does so in minutes on a personal laptop whereas the classical model required hundreds of hours on a high-performance computing cluster. In chapter two, I present a new R package, ForestPlotR,

which provides a toolkit for the exploration and analysis of mapped forest stand data, including an implementation of the regularized regression tree growth model developed in chapter one. Together, the products of these first two chapters have the potential to increase the accessibility of mapped forest stand data, and will hopefully encourage exciting new research directions in this field. Despite the power of big data, it is important to recognize that big ecological datasets typically arise from the combining of many contributed datasets based on a diversity of study systems. In this age of big data, we must remember that each of these contributed datasets represents the result of a well-designed experiment, and such experiments can only be developed by researchers with an intimate knowledge of their system obtained through detailed observation. In chapter three, I demonstrate this critical axis of contemporary ecology by investigating how the climate change-mediated range shifts of a coniferous tree may be moderated by the distributions of their mutualistic fungi. Surprisingly, I find some evidence that high dependence on a mutualism may not slow the rate of range shift. Overall, this dissertation highlights some important considerations that we ecologists must make as our field adapts to its increasing preponderance of data.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I am deeply indebted to many colleagues, friends, family members, and funding agencies, whose support and generosity have been crucial in my completion of this dissertation. I cannot list all these wonderful people here, but the following paragraphs highlight the major contributors to this work.

First, my research has been funded by generous grants from the Stuntz Mycology Fund, Mazamas, the Oregon Mycological Society, and the University of Washington Biology Department. I am extremely grateful for the interest and confidence these institutions have shown in my projects.

I would like to especially thank my advisor, Janneke, for her constant encouragement, her transparent and constructive advice, and her unwavering support for my fluctuating career goals. I would not be nearly the researcher I am today, and certainly wouldn't enjoy doing research as much as I do, if not for Janneke's excellent mentorship. I would also like to thank my committee members: Joe, Brian, and Ben. Joe has instilled in me a fascination with fungi that will stay with me forever and his knack for always asking the most difficult questions has improved my critical thinking skills to no end. I will sorely miss our mycological discussions and the many philosophical tangents they took. Brian's unstoppable enthusiasm for forest ecology has inspired many of my research ideas, and I will always remember his community ecology class where he taught us how to infer the successional history of a forest just by looking around. Ben is the most encouraging teacher and person that I know and I cannot thank him enough for his thoughtful advice in my first year of graduate school.

I have also been lucky enough to form collaborations with an amazing set of people who have shaped my projects and my career aspirations. I am especially grateful for my closest collaborator, Ariel, who inspired me to pursue a career in data science and has been instrumental in the design and completion of two of my dissertation chapters. Another irreplaceable collaborator is Olivia Kosterlitz, who has patiently taught me so much about lab work and how to pick myself up when an experiment fails. I also thank Lauren Dorsch, whose inquisitive nature and always positive attitude were critical in helping us overcome the many challenges that arose in our fungal ecology project.

I am so thankful for my friends who have always been there to share in the successes and frustrations of graduate school, and to remind me that there is life outside of work. In particular I want to thank Aji John, Ali Foroozani, David Villalobos, Elli Theobald, Katie Dickinson, Kavya Pradhan, Leonard Jones, Meera Lee Sethi, Rubén Manzanedo, Sujay Balebail, and Tony Cannistra.

Finally I thank my family, whose unending support underlies all my achievements.

Chapter 1. Regularized Regression: A New Tool for Investigating and Predicting Tree Growth

Citation: Graham, S.I., Rokem, A., Fortunel, C., Kraft, N.J.B. & Hille Ris Lambers, J. (2021). Regularized regression: a new tool for investigating and predicting tree growth. Forests, 12, 1283.

Abstract

Neighborhood models have allowed us to test many hypotheses regarding the drivers of variation in tree growth, but require considerable computation due to the many empirically supported non-linear relationships they include. Regularized regression represents a far more efficient neighborhood modeling method, but it is unclear whether such an ecologically unrealistic model can provide accurate insights on tree growth. Rapid computation is becoming increasingly important as ecological datasets grow in size, and may be essential when using neighborhood models to predict tree growth beyond sample plots or into the future. We built a novel regularized regression model of tree growth and investigated whether it reached the same conclusions as a commonly used neighborhood model, regarding hypotheses of how tree growth is influenced by the species identity of neighboring trees. We also evaluated the ability of both models to interpolate the growth of trees not included in the model fitting dataset. Our regularized regression model replicated most of the classical model's inferences in a fraction of the time without using high-performance computing resources. We found that both methods could interpolate out-of-sample tree growth, but the method making the most accurate predictions varied among focal species. Regularized regression is particularly efficient for comparing hypotheses because it automates the process of model selection and can handle correlated explanatory variables. This feature means that regularized regression could also be used to select among potential explanatory variables (e.g., climate variables) and thereby streamline the development of a classical neighborhood model. Both regularized regression and classical methods can interpolate out-of-sample tree growth, but future research must determine whether predictions can be extrapolated to trees experiencing novel conditions. Overall, we conclude that

regularized regression methods can complement classical methods in the investigation of tree growth drivers and represent a valuable tool for advancing this field toward prediction.

1.1. Introduction

Forest plots where the locations of all trees are mapped contain invaluable information for revealing processes of community assembly and dynamics (Condit, 1995; Wiegand et al., 2017). Neighborhood analyses are one particularly common use of such data, and involve modeling some metric of tree performance (usually growth rate) as a function of the species identities, sizes, and other aspects of neighboring trees (Canham & Uriarte, 2006). These neighborhood models are typically used for inferential projects, where several model structures that represent different hypotheses of a process of interest (e.g., competition) are fitted to the same dataset, and model selection is used to determine the hypothesis with most support (Uriarte et al., 2004). This approach has allowed many interesting questions to be addressed, including: How are competitive interactions influenced by environmental variables (Baribault et al., 2012)?; How might tree performance respond to climate change (Buechling et al., 2017)?; How are competitive interactions moderated by niche similarities and hierarchies (Fortunel et al., 2016; Lasky et al., 2014)?

Current neighborhood models are structured around a set of empirically supported patterns of tree performance, but this ecological realism has some drawbacks. For example, tree growth is typically modeled as a non-linear function of tree size and the degree of crowding by neighboring trees (Uriarte et al., 2004). These non-linear relationships increase the number of parameters and the opportunity for local optima, resulting in a slow and computationally intensive model fitting process (for a linearized version, see: Fortunel et al., 2018; Kunstler et al., 2016). In addition, the need to fit a separate model for each hypothesis to be tested further increases computation time and thereby limits the number of hypotheses compared; a potential problem when considering the many ways in which processes such as competition can be modeled (Wiegand et al., 2017). Another consequence of ecological realism is that the models representing different hypotheses tend not to be fully nested, meaning that information theoretic approaches such as Akaike's information criterion (AIC) must be used for model selection. This is not

ideal because AIC penalizes model complexity and can thereby lead to an overly simple final model and bias the conclusions drawn (Tredennick et al., 2021).

Regularized regression is an alternative method for neighborhood analysis that avoids many of these drawbacks. The aim of regularization is the same as AIC model selection; to trade off the antithetical aims of penalizing model complexity and accurately fitting the training data. However, in regularized regression, the strength of this trade-off is determined by a regularization parameter, which is estimated through cross-validation. We focus specifically on Least Absolute Shrinkage and Selection Operator (LASSO) regularized regression (Tibshirani, 1996), which penalizes model complexity by shrinking coefficients of unhelpful covariates to zero and thereby conducts variable selection autonomously. The main benefit of regularized regression is speed; as a linear model, it can be fitted very quickly, and autonomous variable selection removes the requirement of fitting multiple models representing different hypotheses. Moreover, regularized regression is highly robust to correlated variables; therefore, it allows a single model to include many manifestations of the same process and drop all but the most influential driver (Tibshirani, 1996). Importantly, due to its penalization of model complexity, regularization can result in an overly simple final model (Tredennick et al., 2021), but it is unclear whether it leads to more bias than AIC. The main drawback of regularized regression is that it requires a linear modeling framework, which may not accurately approximate the inherently non-linear effects on tree growth (Buechling et al., 2017; Gea-Izquierdo et al., 2013). Overall, regularized regression offers a distinct set of strengths and weaknesses relative to classical neighborhood models.

Another potential use of neighborhood analyses is to predict tree growth beyond the spatial or temporal limits of training data. To the best of our knowledge, neighborhood models have not been used for out-of-sample prediction, but this is likely to change as growing emphasis on dataset publication (Davies et al., 2021; Mahli et al., 2021) and new software tools (Stanke et al., 2020) increases the availability of mapped forest plot data. Moreover, accurate predictions of tree growth are highly desirable because they could provide insight to current spatial patterns in carbon storage and the effects of land use and climate change

on timber production and global carbon cycling. It is important to recognize two levels of prediction: (1) interpolation of observations experiencing conditions contained in the training data (e.g., of trees experiencing environmental conditions similar to those of trees used for model fitting); and (2) extrapolation to observations experiencing conditions not present in the training data (e.g., novel climates). However, neither the interpolation nor extrapolation ability of neighborhood models has been tested.

Both classical neighborhood models and regularized regression may be well-suited to prediction, but it is unclear which method will be the most accurate. The biggest pitfall in predictive modeling is applying an overly complex model that fits the sample data well, but does not generalize to other data because it includes relationships that are specific to the sample. Both AIC and regularization are designed to avoid this overfitting problem by penalizing model complexity, and therefore classical and regularized regression neighborhood models may produce accurate out-of-sample predictions. The cross-validation model selection approach used to select the regularization parameter also guards against overfitting and is growing in popularity in ecology (Dormann et al., 2018; Mac Nally et al., 2018). However, cross-validation has been shown to be asymptotically equivalent to AIC (Stone, 1977) and also requires models to be fitted to many subsets of the data. To ensure that resources are not wasted on applying cross-validation to classical neighborhood models, it is timely to empirically demonstrate the equivalency between AIC and cross-validation model selection.

In this study, we built a novel regularized regression model of tree growth and used it to test hypotheses regarding how tree growth is influenced by the presence and species identity of neighboring trees. This is a particularly interesting test case for a new modeling method because variation in the abundances of different neighbor species leads to an inherently unbalanced sample, which is likely to create challenges in model fitting. To evaluate the accuracy of our model's inferences, we compared them with those of a commonly used neighborhood model (Uriarte et al., 2004; hereafter, likelihood model). In addition, we investigated the ability of both neighborhood models (i.e., regularized regression and likelihood) to

interpolate the growth of trees not used in model fitting. We also evaluated the predictive performance of a cross-validated likelihood model to demonstrate the equivalency between AIC and cross-validation model selection (Stone, 1977). Overall, we found that regularized regression makes similar inferences to the likelihood model, but that the neighborhood model with the most accurate out-of-sample predictions varies between focal tree species.

1.2. Materials and Methods

1.2.1. *Tree Growth Data*

The data used in this study came from the mature and old-growth conifer forests of Mount Rainier National Park, WA, USA. Mount Rainier is a 4392 m high volcano and covers a large climatic gradient. Increasing elevation is associated with decreasing temperatures and increasing precipitation, although precipitation is considerably reduced on the eastern side of the volcano due to a rain shadow effect. The region experiences a temperate maritime climate with warm, dry summers and cold, wet winters.

We used data collected in 15 forest plots established in 1977 and 1978 as part of the Pacific Northwest Permanent Sample Plot Program (Franklin et al., 2021). These plots were intentionally located to capture the diversity of climatic conditions on Mount Rainier, and therefore range in elevation from 581 to 1492 m. All plots are 1 ha (100 × 100 m) in size and, at the time of their establishment, all trees with a diameter at breast height (1.37 m above ground level; hereafter DBH) ≥ 15 cm were tagged, identified to species, mapped on a coordinate grid, and had their DBH recorded. Between 25% and 100% of the area of each plot was also designated as a detailed plot where data were collected on all trees with a DBH ≥ 5 cm. Approximately every five years, all plots are revisited to tag new trees meeting the minimum size threshold, document tree mortality, and to re-measure the size of tagged trees, with the most recent census occurring in 2017.

We calculated average annual growth for each tree as the difference in DBH between its earliest and most recent measurement divided by the number of years elapsed between those measurements. The slow

growth rates of trees in this harsh high-elevation environment meant that measurement inaccuracies sometimes resulted in biologically impossible negative growth rates; all such trees were excluded from our analysis (1.6% of focal trees). The smaller trees recorded only in the detailed plots (5–15 cm DBH) were included in our analyses as focal trees but excluded as neighbors that could influence the growth of other focal trees to prevent systematic bias in neighbor interactions between detailed and non-detailed areas of the plots. Of the 17 tree species included in the dataset, we modeled growth for only the six species represented by at least 100 individuals: *Abies amabilis*, *Callitropsis nootkatensis*, *Pseudotsuga menziesii*, *Thuja plicata*, *Tsuga heterophylla*, and *Tsuga mertensiana* (hereafter: ABAM, CANO, PSME, THPL, TSHE, TSME, respectively; see Table S1.1 for full species list and Table S1.2 for a summary of how focal species were distributed across sample plots).

For our neighborhood models, we considered all trees growing within 15 m of a focal tree to be that focal tree's neighbors. This neighborhood size is comparable to those used in other studies (Uriarte et al., 2004; Fortunel et al., 2016) and was found through our own exploratory analyses to result in the best training data fits. To avoid edge effects, all focal trees within 15 m of a forest plot boundary were excluded from our analysis. Each of the remaining focal trees was assigned to one of four test datasets at random, such that 25% of the focal trees in each plot were placed in each test set. We then defined a corresponding training dataset for each test set such that the first training set consisted of all focal trees not in the first test set (75% of all focal trees). All models were fitted to each of the four training sets to assess the robustness of conclusions; test sets were used to evaluate the predictive skill. As in previous studies, we also created separate models for each of our focal species because model parameter values are expected to differ greatly between species.

The data and code underlying all presented analyses are available on Zenodo at DOI: 10.5281/zenodo.5512791 (Graham et al., 2021).

1.2.2. *Likelihood Model*

The likelihood model (Uriarte et al., 2004; Canham et al., 2006) has the following generalized formula:

$$g = g_{max} \times \delta_t \times \gamma_p \times \omega_t, \quad (1.1)$$

where g is the predicted growth, g_{max} is an estimated maximum potential growth rate in the absence of neighbors, δ is a size effect, γ is a climate effect, and ω is a crowding effect, with subscripts indicating whether effects vary between focal trees (t) or plots (p). The size, climate and crowding effects can take any value between 0 and 1; therefore, growth predictions can take any value between 0 and g_{max}.

The size effect (δ) accounts for the expectation that trees have an optimal size at which maximum growth occurs and is modeled as a lognormal distribution:

$$\delta_t = \exp\left(-\frac{1}{2}\left(\frac{\log(DBH_t/X_0)}{X_b}\right)^2\right), \quad (1.2)$$

Parameter X_0 specifies the DBH at which maximum growth occurs, and parameter X_b determines the width of the lognormal distribution. This formulation is highly flexible, allowing the relationship between focal growth and size to be monotonically increasing, monotonically decreasing, or non-monotonic.

The climate effect (γ) accounts for the expectation that growth rates will differ among the plots due to their different climatic conditions. Although there are many climatic variables that differ dramatically along the elevational gradient on which our plots are situated, these variables are strongly correlated (Ettinger et al., 2011), and potential evapotranspiration (PET) is informative of growth rates of our focal species in these plots (Ford et al., 2017). Consequently, we used PET as the sole abiotic variable in our models and calculated the average annual PET for each plot from the time of plot establishment up until the most recent tree measurements, following the protocol outlined in (Ford et al., 2017). The climate effect (γ) is modeled as a Gaussian distribution:

$$\gamma_p = \exp\left(-\frac{1}{2}\left(\frac{\mathrm{PET}_p - pet_a}{pet_b}\right)^2\right), \quad (1.3)$$

where PET_p represents the average annual PET of plot p (where the focal tree resides), pet_a specifies the PET at which maximum growth occurs, and pet_b determines the width of the Gaussian distribution. As

with the size effect, this flexible structure allows the relationship between focal growth and PET to monotonically increase, monotonically decrease, or be non-monotonic.

To incorporate the effects of neighbors on focal tree growth, a neighborhood crowding index (NCI) was calculated for tree t as:

$$NCI_t = \sum_{i=1}^{S} \sum_{j=1}^{n_i} \frac{{DBH_{ij}}^{\alpha}}{{Distance_{ij}}^{\beta}} \times \lambda_i, \quad (1.4)$$

where S is the number of neighbor species and n_i is the number of trees of species i in focal tree t's neighborhood. This formula reflects the expectation that a neighbor's influence on focal tree growth increases with its size but decreases with its distance from the focal, and the estimated parameters α and β allow these relationships to be non-linear. The effect of neighbor size and distance is also multiplied by an estimated interaction coefficient (λ_i), which takes a value between 0 and 1 and represents the effect of neighbors of species i on the growth of focal trees of the species being modeled.

The crowding effect (ω) is calculated as a negative exponential function of NCI:

$$\omega_t = \exp(-C \times NCI_t), \quad (1.5)$$

where C is an estimated parameter that modulates the growth response of trees to varying NCI values.

To make inferences regarding the effects of neighbors on focal growth, we fitted four variations of the likelihood model for each focal species: (1) no interactions – crowding effect (ω) excluded; (2) equivalent interactions – no λ parameters included (quantitatively equivalent to a single λ with value 1); (3) conspecific vs. heterospecific interactions – two λ parameters, one for conspecific neighbors (λ_{con}) and another for heterospecific neighbors (λ_{het}); and (4) species-specific interactions – estimated λ_i for each neighbor species i. In the species-specific interaction models, rare neighbor species were grouped under a single λ_{other} parameter. Rare neighbor species (<5% of neighbors in each focal species × training set combination) were defined as those that appeared as neighbors of the focal species fewer than 100 times, when averaged across the four training sets. We elected to use an average instead of specifying rare

neighbor species separately for each training set to ensure that the fitted λ_{other} parameters could be compared across training sets. The four model structures for each focal species × training set combination were compared using Akaike's information criterion corrected for a low sample size (AICc).

Parameter values were estimated using the simulated annealing algorithm implemented through the *optim* function in the base library of R 4.0.2 (R Core Team, 2020). The optimizations were facilitated through the use of advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system at the University of Washington.

1.2.3. *Regularized Regression Model*

In our regularized regression model, focal tree growth was modeled as a linear function of: the species identity, size and proximity of neighbors; PET; and the densities of each neighbor species, and all species combined, in the neighborhood. Rare neighbor species (characterized in the same way as for the likelihood model) were assigned a species identity of "other", and the density of "other" was also included in the model. To estimate the effects of neighbor identity, size and proximity in a linear modeling framework, each focal tree-neighbor interaction was treated as an independent observation. This resulted in a design matrix where each focal tree occupied n rows, with n being the number of neighbors in its neighborhood. The growth rate, PET and densities were necessarily identical across all n rows corresponding to the same focal tree. This design matrix structure resulted in n growth predictions of each focal tree based on each of its n interactions and we used the arithmetic average of these predictions as the final prediction of focal growth.

To meet the assumptions of linear regression, the growth rate data were transformed to approximate a normal distribution as follows:

$$g = \sqrt{\frac{average\ annual\ diameter\ growth}{initial\ DBH}}, \tag{1.6}$$

This transformation has the additional benefit of partially accounting for the non-linear relationship between focal size and focal growth included in the likelihood model; it allows a saturating but always monotonic relationship between growth rate and tree size.

We fitted the regularized regression models using the *cv.glmnet* function of the glmnet R package (Friedman et al., 2010). This function estimates parameter values through a stochastic cyclical coordinate descent algorithm using a set of values for the regularization parameter. It then uses 10-fold cross-validation to evaluate the models fitted with different regularization parameter values, reporting the mean square error (MSE) for each. Of the multiple output models, glmnet indicates the one with the highest regularization value (strongest regularization) that resulted in an MSE within one standard error of the model with the lowest MSE; we used this model for interpretation. In addition, the rapid fitting of the regularized regression models allowed us to fit 100 models for each focal species by training set combination to evaluate how consistent the findings of this model are in the face of the stochastic fitting process. Of these 100 models, the one with the lowest MSE was used for evaluating model fit to the training data and predictive performance. The model fitting procedure implemented by glmnet is rapid; therefore, it was conducted on a personal laptop.

1.2.4. *Comparing Inferential Performance*

To determine whether our regularized regression model could replicate the inferences of the likelihood model, we compared the conclusions each of the models would have led us to for four commonly asked questions regarding the impact of neighbors on tree growth. Separately for each focal species, we asked: (1) Is focal growth influenced by neighboring trees?; (2) Is focal growth influenced by neighbor species identity?; (3) Is focal growth higher in the presence of conspecific or heterospecific neighbors?; and (4) Which neighbor species are associated with the highest/lowest focal growth? For each modeling approach, the conditions under which we drew particular conclusions regarding these questions are outlined in Table 1.1.

1.2.5. *Evaluating Predictive Performance*

To investigate the predictive potential of neighborhood models, we evaluated the out-of-sample predictive ability for three different models for each focal species by training set combination: regularized regression, AIC likelihood and CV likelihood. Each of these predictive models was one of the models described in Methods: Inference. For the regularized regression model, we used the model with the lowest MSE (out of the 100 models run). For the AIC and CV likelihood models, we used the models with the lowest AIC and lowest cross-validated MSE, respectively, of the four model structures. To identify the lowest cross-validated MSE, we divided each training set into 10 folds (consistent with regularized regression cross-validation), fitted each of the four likelihood model structures to each possible set of 9 folds, then calculated MSE of the models' predictions of the 10th fold. We averaged the resulting 10 MSE values to obtain the cross-validated MSE of each model structure.

We quantified the out-of-sample prediction (interpolation) ability of each model type applied to each training set by calculating the coefficient of determination (R^2) of the model when applied to its corresponding test set, which was entirely unused in the fitting of that model (see Section 1.2.1 Tree growth data for how training and test sets were defined). This metric measures the proportion of variance around the mean value of the dependent variable explained by the model. The maximum possible value for a coefficient of determination is 1 (all variance explained), but negative values can exist when a model is applied to unseen test data if there is more unexplained variation around model predictions than exists around the mean growth value in the test data.

It is often advised that training and test sets be spatially or temporally separated to prevent overestimates of predictive ability that can result from spatial or temporal autocorrelation. In this study, we present results from spatially overlapping training and test sets because the predictive performance of the models was similar when the spatially separated training and test sets were used. Moreover, although we suspect that spatial and temporal autocorrelation in unmeasured variables which influence tree growth was likely

in our dataset (e.g., soil conditions, pest damage), we do not know the scale of such variation – which means that it is unclear whether spatial or temporal separation would address such autocorrelation.

1.3. Results

1.3.1. *Comparing Inferential Performance*

The regularized regression models generally led to the same qualitative conclusions regarding tree growth hypotheses, as did the AIC likelihood models, but did so in a fraction of the time. In combination, the regularized regression models took <15 min to fit on a personal laptop, whereas the AIC likelihood models took 339 h on the HYAK supercomputer system. The AIC likelihood models were quite consistent across training sets with regard to the best-fitting model structure (Table S1.3) and fitted parameters (Tables S1.4–S1.9; Figures S1.1 and S1.2). The regularized regression models also showed consistency in the variables retained and parameter values (Tables S1.10–S1.15). The two modeling methods agreed that growth in all focal species was influenced by both neighboring trees and their species identity in all training sets, with the one exception of the likelihood model finding PSME growth to be unaffected by neighbor species identity in one training set. There was also overall agreement between methods on whether focal growth was higher in the presence of conspecific or heterospecific neighbors (Table 1.2). With perfect consistency across training sets, both methods found ABAM to grow faster in the presence of conspecific neighbors, but found CANO, TSHE, and TSME to grow faster in the presence of heterospecific neighbors. The regularized regression found conspecific and heterospecific neighbors to be associated with equal focal tree growth rates in all PSME training sets and three of the THPL training sets. Somewhat consistent with this, the conclusions of the AIC likelihood model for PSME and THPL changed the direction across training sets (Table S1.16).

The regularized regression and AIC likelihood methods showed less agreement on which neighbor species were associated with the highest and lowest focal growth (Figures 1.1 and S1.3–S1.7). Some examples of agreement were that ABAM grew quickly in the presence of conspecific neighbors, but

slowly in the presence of THPL and TSME neighbors (Figure 1.1), and TSHE grew slowly in the presence of CANO and PSME neighbors (Figure S1.6). The most common scenario of disagreement was when the AIC likelihood model found a neighbor species to be associated with high or low focal growth, but the regularized regression defaulted to medium (neighbor identity variable dropped from model). The likelihood model sometimes concluded that a neighbor species was associated with low growth of a particular focal species in one training set but high growth of that same focal species in another training set (e.g., the effect of PSME on focal ABAM; Figure 1.1); in contrast, regularized regression was much more consistent across training sets. The only case of direct disagreement was in the effect of ABAM on focal PSME (Figure S1.4).

1.3.2. *Evaluating Predictive Performance*

The modeling method that resulted in the highest out-of-sample predictive skill differed between focal species (Figure 1.2B). The likelihood models (AIC and CV) made more accurate predictions than the regularized regression models for ABAM and THPL, but regularized regression performed best for PSME. All three methods had similar predictive skill for CANO, TSHE and TSME, although only regularized regression maintained some predictive skill (coefficient of determination > 0) for TSME in all training sets. Importantly, the fit to training data (Figure 1.2A) was often considerably higher than the fit to out-of-sample data (Figure 1.2B; e.g., CANO and TSME). The AIC and CV likelihood models had almost identical out-of-sample predictive skill (Table S1.17); however, the CV model required far more computation (881 h on the high-performance computing cluster). The final CV likelihood model often had a different structure from the final AIC likelihood model, but was not consistently more or less complex (Tables S1.3 and S1.18).

1.4. Discussion

We found that our inferential conclusions regarding the effects of neighbors would have been very similar whether using the regularized regression model or the more realistic likelihood model of tree growth. In

contrast, the method that led to the most accurate out-of-sample predictions varied between focal species. The biggest difference between the two methods was in the time taken for model fitting, with the regularized regression taking minutes on a personal computer and the likelihood approach requiring many hours on a computing cluster. In all, we believe the two approaches to each have their advantages, and we encourage further investigation of how novel regularized regression can complement the classical likelihood method.

1.4.1. *Using Regularized Regression for Inference*

The regularized regression model generally reached the same conclusions as the likelihood model regarding hypotheses for how focal tree growth is affected by neighbors. The two modeling approaches were in almost total agreement that the growth of all focal species was influenced by neighbors and their species identities. We have confidence in these conclusions because they align with previous studies which found tree growth to be influenced by: (1) the local neighborhood in these plots (Ford et al., 2017; Ettinger & Hille Ris Lambers, 2013); and (2) neighbor species identity in similar forests of British Columbia (Canham et al., 2004; Coates et al., 2009). The two modeling approaches also agreed that focal trees of ABAM grew more quickly in the presence of conspecifics, but that the opposite was true for CANO, TSHE, and TSME (Table 1.2). Furthermore, regularized regression concluded that the growth of PSME and THPL was similar in the presence of conspecific and heterospecific neighbors, which is consistent with the likelihood model's findings of small and variable signals across training sets for PSME and THPL, respectively (Table S1.16). The positive and negative feedbacks on the growth of ABAM and TSHE, respectively, and the similar effects of conspecific and heterospecific neighbors on THPL, do have some support in the literature (Canham et al., 2004; Coates et al., 2009), but we are not aware of previous assessments for CANO, TSME, or PSME. The agreement between regularized regression and likelihood models was lower when considering which neighbor species were associated with high/low focal growth (Figures 1.1 and S1.3–S1.7). Together, our results demonstrate strong agreement in the inferences drawn from regularized regression and likelihood models, particularly in

broad-scale patterns, and it is unclear whether the generally more complex likelihood models represent a lack of sensitivity of regularized regression or false detection by the likelihood model.

Our novel regularized regression approach may be particularly well-suited to the inferential analysis of large data sets due to its low computational requirements and automation of model selection. By transforming the growth variable to account for its non-linear relationship with focal tree size, we constructed an accurate linear model that could be fitted in minutes on a personal computer. However, it is certainly possible that in datasets where the degree of non-linearity is more extreme, linearizing transformations will be insufficient (Gea-Izquierdo et al., 2013). For example, our growth transformation does not permit a decrease in growth rates at larger tree sizes, an often-documented pattern that was very weak in our dataset. It should be noted that linearization has been applied to classical neighborhood models, thereby reducing their computational requirements (Fortunel et al., 2018; Kunstler et al., 2016), but these models still fit far more slowly than regularized regression and do not automate model selection. In a likelihood modeling framework, a separate model representing each of the several hypotheses to be tested must be fitted independently, greatly increasing the total model-fitting time and constraining possible conclusions to those permitted by comparisons among the finite set of constructed models. This can be problematic due to the many different ways there are to model processes such as competition (Wiegand et al., 2017). In contrast, an almost unlimited number of variables pertaining to different models of competition (e.g., measures of functional and phylogenetic distances) can be included in a single regularized regression model, and those that have the greatest influence will be indicated. This model selection capability suggests that regularized regression could also be used as an exploratory tool prior to developing a likelihood model by, for example, selecting among the many potential climate variables to include.

1.4.2. *Using Neighborhood Models for Prediction*

The absolute and relative abilities of the likelihood and regularized regression methods to predict out-of-sample data are likely to vary among species and systems. Both methods demonstrated some out-of-

sample predictive ability for all focal species in at least some training sets, although this ability was weak and variable for TSME, the focal species with the lowest sample size (Figure 1.2). The likelihood model, whether using AIC or cross-validation for model selection, more accurately predicted the growth of ABAM and THPL, whereas the regularized regression made better predictions for PSME. The high predictive skill of the regularized regression model for PSME may be due to the apparently small effect of neighbors on this species' growth, indicated by the small fitted value of parameter C in the likelihood model (Table S1.6). This suggests that PSME growth can be predicted using only the focal tree size and PET, which may be reasonably approximated with linear relationships (Figures S1.1 and S1.2). Similarly, the higher predictive ability of the likelihood model for ABAM could be a result of the highly non-linear PET effect observed in this species (Figure S1.2). We also found the regularized regression method to show more consistent predictive ability across training sets, as evidenced by it being the only method to maintain predictive skill across all training sets for CANO (Figure 1.2B). As was predicted from the proven equivalency of AIC and cross-validation (Stone 1977), the predictive performance of the likelihood model did not increase when the more computationally intensive cross-validation technique was used for model selection.

Further research is certainly needed before neighborhood models can be used for prediction purposes, but our results do provide some important lessons. We have shown that both the likelihood and regularized regression models can usually interpolate the growth rates of trees experiencing similar conditions to those in the training set. However, a test of whether these models' predictions can be extrapolated to trees experiencing different conditions, such as novel climates, is still lacking, and we encourage research in this area. We also learned that the inclusion of neighbor interactions generally improves the accuracy of tree growth predictions, as evidenced by the results of our AIC and cross-validation model selection that are designed for predictive modeling. Additionally, our finding of poor out-of-sample predictive ability for some focal species (e.g., TSME) reinforces the practice of always evaluating predictive models using a test set (Tredennick et al., 2021). Although setting aside 20% to 25% of a hard-earned dataset for testing

can be unsettling, we argue that this practice is critical, especially where results may influence stakeholder decision-making. Moreover, ecological datasets are quickly growing larger and more accessible, thereby enabling us to alter best practices as our ability to address many important questions in the ecological field becomes less limited by data. Another lesson learned is that the most appropriate neighborhood modeling method will vary among species, with potentially large effects on predictive performance (e.g., PSME; Figure 1.2B). We advise consideration of whether linearizing transformations are appropriate for a given dataset before deciding whether to use regularized regression vs. likelihood approaches.

Regularized regression models may be particularly useful for predictive projects due to their rapid fitting and ability to include many correlated explanatory variables. Most predictive models are intended for application over large spatial areas, and consequently will be trained on large datasets. As a result, models that can be fitted quickly may be necessary, and regularized regression has a clear advantage over likelihood approaches in this regard, even when likelihood models are linearized (Fortunel et al. 2018). We also argue that the full potential of regularized regression models is far from realized. Our regularized regression models accounted only for the non-linear relationship between growth and focal size, but it may be possible to use other transformations to accommodate other complex relationships. For example, the influence of a neighboring tree is expected to vary non-linearly with that neighbor's DBH—this could be incorporated in a linear framework by using: (1) $\sqrt{DBH}$; (2) DBH^2; (3) only neighbors with DBH > focal DBH, etc. In a likelihood framework, a separate model would need to be fitted for each of these neighbor DBH transformations, whereas a regularized regression could incorporate all of them at once and indicate the most informative, which could then optionally be used to design a likelihood model.

1.4.3. *Conclusions*

We have developed a regularized regression model of neighborhood-dependent tree growth that can replicate the ecological inferences of a classical likelihood model in a fraction of the time. Regularization is particularly efficient for inferential projects because it automates the process of model selection and can

handle correlated explanatory variables. This feature means that our regularized regression model could also be used to select among potential explanatory variables (e.g., climate variables) and thereby streamline the development of a classical likelihood model. We encourage the investigation of regularization as a tool for modeling tree growth and other processes that have many potential covariates, such as seedling survival (Metz et al., 2010) and mature tree mortality (Buonanduci et al., 2020).

We have also shown that neighborhood models, including regularized regression, can provide accurate growth predictions of trees not used in model fitting. However, we only tested the model's predictive skill on trees that experienced similar conditions to those used in model fitting. Future research should investigate whether the findings of neighborhood models can be extrapolated to trees experiencing conditions absent from the training set (e.g., novel climates). Although the optimal neighborhood modeling approach for prediction will vary among species and systems, we believe that our regularized regression has great potential due to its rapid fitting and ability to include many explanatory variables that represent different models of complex processes such as competition. Overall, we found that regularized regression and likelihood approaches are complementary to better understand the drivers of tree growth, and suggest that regularization will be a valuable tool for advancing the field of tree growth modeling toward prediction.

1.5 Data Availability Statement

The data presented in this study are openly available on Zenodo at DOI: 10.5281/zenodo.5512791 (Graham et al., 2021).

1.6 Acknowledgments

We thank the UW eScience Incubator program and the maintainers of the UW Hyak supercomputer system. We also thank Joe Ammirati, Tony Cannistra, Aji John, Rubén Manzanedo, Kavya Pradhan, and Meera Lee Sethi for feedback on project development. Tree growth data were provided courtesy of the Pacific Northwest Permanent Sample Plot Program, in partnership with the HJ Andrews Experimental Forest and Long Term Ecological Research (LTER) program, which are administered cooperatively by the USDA Forest Service Pacific Northwest Research Station, Oregon State University, and the Willamette National Forest.

1.7 References

Baribault, T. W., Kobe, R. K., Finley, A. O. (2012). Tropical Tree Growth Is Correlated with Soil Phosphorus, Potassium, and Calcium, Though Not for Legumes. *Ecological Monographs*, *82*, 189–203.

Buechling, A., Martin, P. H., Canham, C. D. (2017). Climate and Competition Effects on Tree Growth in Rocky Mountain Forests. *Journal of Ecology*, *105*, 1636–1647.

Buonanduci, M. S., Morris, J. E., Agne, M. C., Harvey, B. J. (2020). Neighborhood Context Mediates Probability of Host Tree Mortality in a Severe Bark Beetle Outbreak. *Ecosphere*, *11*, e03236.

Canham, C. D., LePage, P. T., Coates, K. D. (2004). A Neighborhood Analysis of Canopy Tree Competition: Effects of Shading versus Crowding. *Canadian Journal of Forest Research*, *34*, 11.

Canham, C. D., Papaik, M. J., Uriarte, M., McWilliams, W. H., Jenkins, J. C., Twery, M. J. (2006). Neighborhood Analyses of Canopy Tree Competition along Environmental Gradients in New England Forests. *Ecological Applications*, *16*, 540–554.

Canham, C. D., Uriarte, M. (2006). Analysis of Neighborhood Dynamics of Forest Ecosystems Using Likelihood Methods and Modeling. *Ecological Applications*, *16*, 62–73.

Coates, K. D., Canham, C. D., LePage, P. T. (2009). Above- versus below-Ground Competitive Effects and Responses of a Guild of Temperate Tree Species. *Journal of Ecology*, *97*, 118–130.

Condit, R. (1995). Research in Large, Long-Term Tropical Forest Plots. *Trends in Ecology & Evolution*, *10*, 18–22.

Davies, S. J., Abiem, I., Abu Salim, K., Aguilar, S., Allen, D., Alonso, A., Anderson-Teixeira, K., Andrade, A., Arellano, G., Ashton, P. S., et al. (2021). ForestGEO: Understanding Forest Diversity and Dynamics through a Global Observatory Network. *Biological Conservation*, *253*, 108907.

Dormann, C. F., Calabrese, J. M., Guillera-Arroita, G., Matechou, E., Bahn, V., Bartoń, K., Beale, C. M., Ciuti, S., Elith, J., Gerstner, K., et al. (2018). Model Averaging in Ecology: A Review of Bayesian, Information-theoretic, and Tactical Approaches for Predictive Inference. *Ecological Monographs*, *88*, 485–504.

Ettinger, A. K., Ford, K. R., HilleRisLambers, J. (2011). Climate Determines Upper, but Not Lower, Altitudinal Range Limits of Pacific Northwest Conifers. *Ecology*, *92*, 1323–1331.

Ettinger, A. K., HilleRisLambers, J. (2013). Climate Isn't Everything: Competitive Interactions and Variation by Life Stage Will Also Affect Range Shifts in a Warming World. *American Journal of Botany*, *100*, 1344–1355.

Ford, K. R., Breckheimer, I. K., Franklin, J. F., Freund, J.A., Kroiss, S. J., Larson, A. J., Theobald, E. J., HilleRisLambers, J. (2017). Competition Alters Tree Growth Responses to Climate at Individual and Stand Scales. *Canadian Journal of Forest Research*, *47*, 53–62.

Fortunel, C., Lasky, J. R., Uriarte, M., Valencia, R., Wright, S. J., Garwood, N. C., Kraft, N. J. B. (2018). Topography and Neighborhood Crowding Can Interact to Shape Species Growth and Distribution in a Diverse Amazonian Forest. *Ecology*, *99*, 2272–2283.

Fortunel, C., Valencia, R., Wright, S. J., Garwood, N. C., Kraft, N. J. B. (2016). Functional Trait Differences Influence Neighbourhood Interactions in a Hyperdiverse Amazonian Forest. *Ecology Letters*, *19*, 1062–1070.

Franklin, J. F., Bell, D., Shaw, D. C. Long-Term Growth, Mortality and Regeneration of Trees in Permanent Vegetation Plots in the Pacific Northwest, 1910 to Present ver 18. Environmental Data Initiative. Available online: https://doi.org/10.6073/pasta/45a1f16b3d8ecd0585a2d2e115c07d41 (accessed on 14 September 2021).

Friedman, J., Hastie, T., Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, *33*, 1.

Gea-Izquierdo, G., Fernández-de-Uña, L., Cañellas, I. (2013). Growth Projections Reveal Local Vulnerability of Mediterranean Oaks with Rising Temperatures. *Forest Ecology & Management*, *305*, 282–293.

Graham, S. I., Rokem, A., Fortunel, C., Kraft, N. J. B., Hille Ris Lambers, J. (2021). sgraham9319/TreeGrowth: Forests manuscript code release; Zenodo; DOI: 10.5281/zenodo.5512791.

Kunstler, G., Falster, D., Coomes, D. A., Hui, F., Kooyman, R. M., Laughlin, D. C., Poorter, L., Vanderwel, M., Vieilledent, G., Wright, S. J., et al. (2016). Plant Functional Traits Have Globally Consistent Effects on Competition. *Nature*, *529*, 204–207.

Lasky, J. R., Uriarte, M., Boukili, V. K., Chazdon, R. L. (2014). Trait-Mediated Assembly Processes Predict Successional Changes in Community Diversity of Tropical Forests. *Proceedings of the National Academy of Sciences USA*, *111*, 5616–5621.

Mac Nally, R., Duncan, R. P., Thomson, J. R., Yen, J. D. L. (2018). Model Selection Using Information Criteria, but Is the "Best" Model Any Good? *Journal of Applied Ecology*, *55*, 1441–1444.

Malhi, Y., Girardin, C., Metcalfe, D. B., Doughty, C. E., Aragão, L. E. O. C., Rifai, S. W., Oliveras, I., Shenkin, A., Aguirre-Gutiérrez, J., Dahlsjö, C. A. L., et al. (2021). The Global Ecosystems Monitoring Network: Monitoring Ecosystem Productivity and Carbon Cycling across the Tropics. *Biological Conservation*, *253*, 108889.

Metz, M. R., Sousa, W. P., Valencia, R. (2010). Widespread Density-Dependent Seedling Mortality Promotes Species Coexistence in a Highly Diverse Amazonian Rain Forest. *Ecology*, *91*, 11.

R Core Team. R: A Language and Environment for Statistical Computing; R Foundation for Statistical Computing: Vienna, Austria, 2020.

Stanke, H., Finley, A. O., Weed, A. S., Walters, B. F., Domke, G. M. (2020). RFIA: An R Package for Estimation of Forest Attributes with the US Forest Inventory and Analysis Database. *Environmental Modeling & Software*, *127*, 104664.

Stone, M. (1977). An Asymptotic Equivalence of Choice of Model by Cross-Validation and Akaike's Criterion. *Journal of the Royal Statistical Society Series B (Methodological)*, *39*, 44–47.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society Series B (Methodological)*, *58*, 267–288.

Tredennick, A. T., Hooker, G., Ellner, S. P., Adler, P. B. (2021). A Practical Guide to Selecting Models for Exploration, Inference, and Prediction in Ecology. *Ecology*, *102*, e03336.

Uriarte, M., Canham, C. D., Thompson, J., Zimmerman, J. K. (2004). A Neighborhood Analysis of Tree Growth and Survival in a Hurricane-Driven Tropical Forest. *Ecological Monographs*, *74*, 591–614.

Wiegand, T., Uriarte, M., Kraft, N. J. B., Shen, G., Wang, X., He, F. (2017). Spatially Explicit Metrics of Species Diversity, Functional Diversity, and Phylogenetic Diversity: Insights into Plant Community Assembly Processes. *Annual Reviews in Ecology, Evolution, & Systematics*, *48*, 329–351.

1.8 Figures and Tables

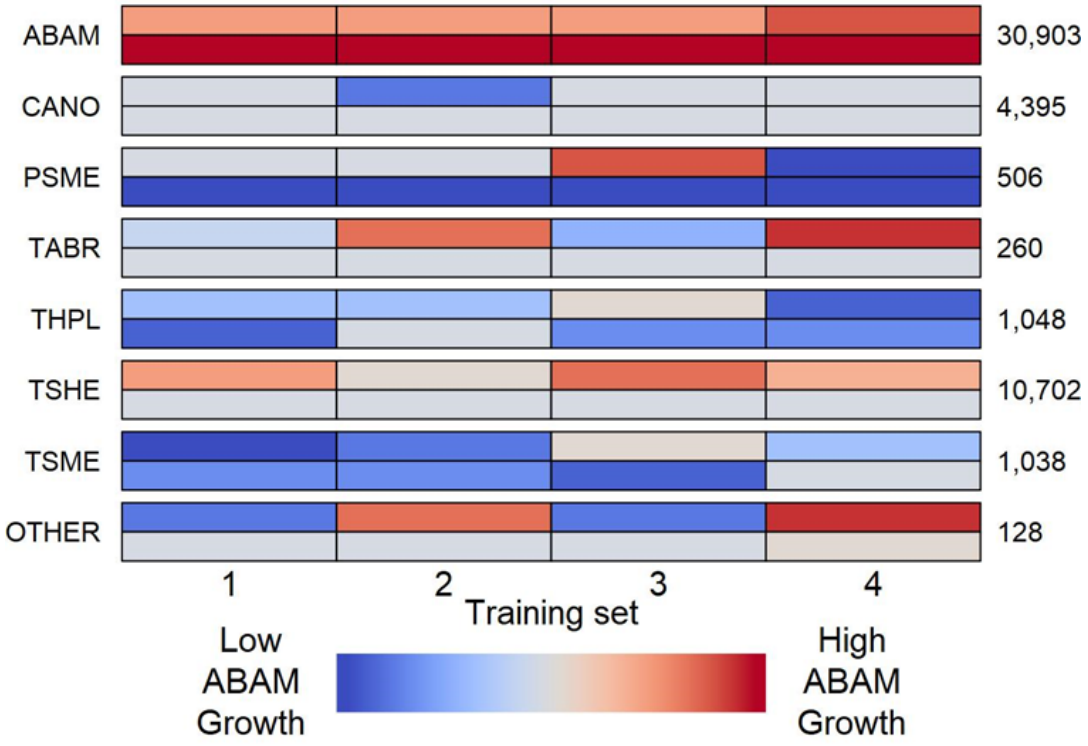

Figure 1.1. Which neighbor species are associated with the highest/lowest growth of ABAM focals? For each neighbor species, there are two rows of colored bars. The top row of bars shows the likelihood model results, and the bottom row shows the regularized regression model results. Each row of bars is divided into four sections, to show the results according to the models fitted to each of the four training sets. The color of the bars indicates the growth rate of ABAM in the presence of the neighbor species that row represents (see inset legend). The numbers on the right of the figure indicate the number of neighbors of each species averaged across training sets.

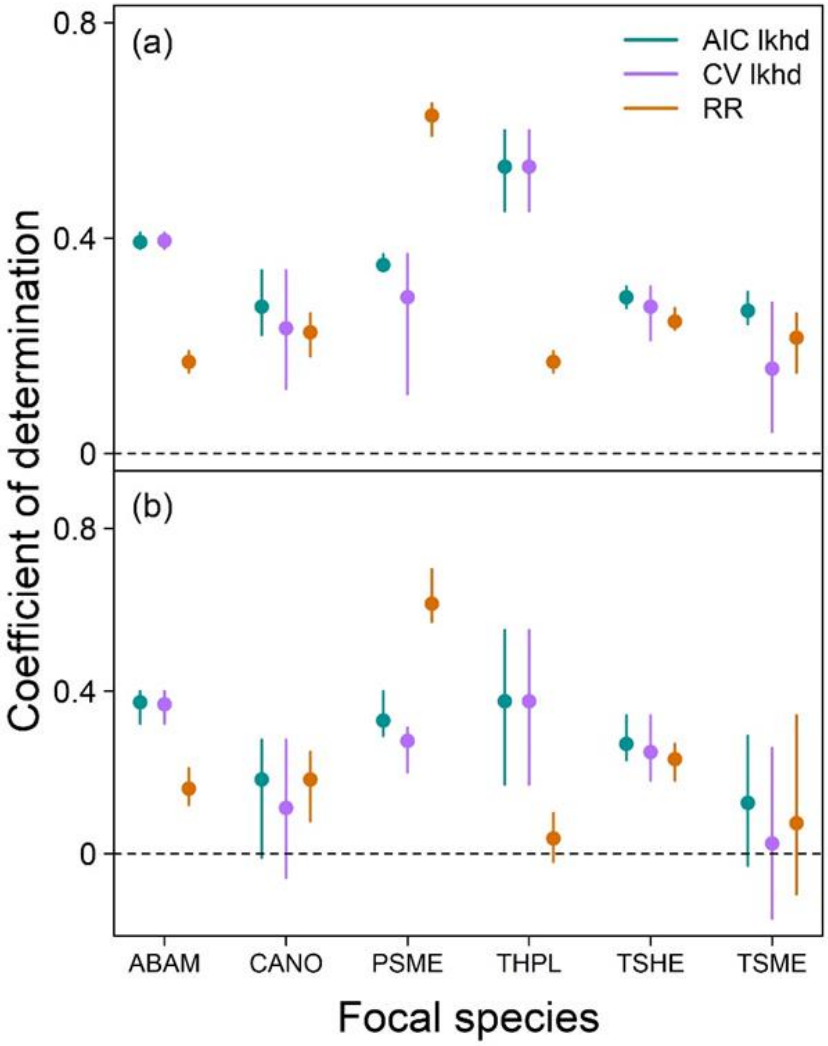

Figure 1.2. Training fit (a) and out-of-sample predictive skill (b) of the likelihood and regularized regression models. Likelihood models using AIC and cross-validation for model selection are shown in green and purple, respectively. Regularized regression models are shown in orange. Points and error bars represent the mean and range of coefficients of determination across the four training sets. Raw values are provided in Table S1.17.

Table 1.1. Conditions used to draw conclusions regarding inferential questions in the likelihood and regularized regression models.

Conclusion	Condition	
	Likelihood	**Regularized Regression** [1]
Focal growth is influenced by neighboring trees	Best model is: equivalent, conspecific vs. heterospecific, or species-specific	At least one species identity, size, proximity, or density variable retained
Focal growth is influenced by neighbor species identity	Best model is: conspecific vs. heterospecific or species-specific	At least one species identity or species-specific density variable retained
Focal growth is higher in the presence of conspecifics	In the best conspecific vs. heterospecific model, $\lambda_{het} - \lambda_{con} > 0$	Coefficient of: neighbor species = focal species and/or focal species density > 0 [2]
Neighbor species X is associated with: (1) high; (2) medium; (3) low focal growth relative to the average neighbor	In the best species-specific model: (1) $\lambda_X > 0.66$, (2) $0.33 \leq \lambda_X \leq 0.66$, (3) $\lambda_X < 0.33$	Coefficient of neighbor species = X is: (1) positive, (2) dropped from model, (3) negative [3]

[1] Conclusions were drawn separately for each of the 100 regularized regression models run for each focal species × training set combination. [2] If coefficients of neighbor species = focal species and focal species density had opposite signs, we concluded that focal growth was unaffected by whether neighbors were conspecific or heterospecific in that particular model run. [3] To enable comparison with λ_X in the likelihood model, we calculated the number of the 100 regularized regression models where the coefficient of neighbor species = X was positive minus the number where this coefficient was negative to obtain a number between −100 and +100, then re-scaled these values to a range of 0–1.

Table 1.2. Is focal growth higher in the presence of conspecific or heterospecific neighbors? Higher growth in the presence of conspecifics represents positive feedback on growth and is indicated with '+'. Negative feedback and the absence of feedback are indicated with '−' and '0', respectively. NA values for regularized regression indicate that the model did not find growth to be substantially higher in the presence of conspecifics or heterospecifics. See Table S1.16 for numerical outputs.

Focal	**Likelihood**				**Regularized Regression**			
Species	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
ABAM	+	+	+	+	+	+	+	+
CANO	−	−	−	−	−	−	−	−
PSME	+	−	−	−	NA	NA	NA	NA
THPL	−	−	0	−	NA	−	NA	NA
TSHE	−	−	−	−	−	−	−	−
TSME	−	−	−	−	−	−	−	−

1.9 Supplementary Materials

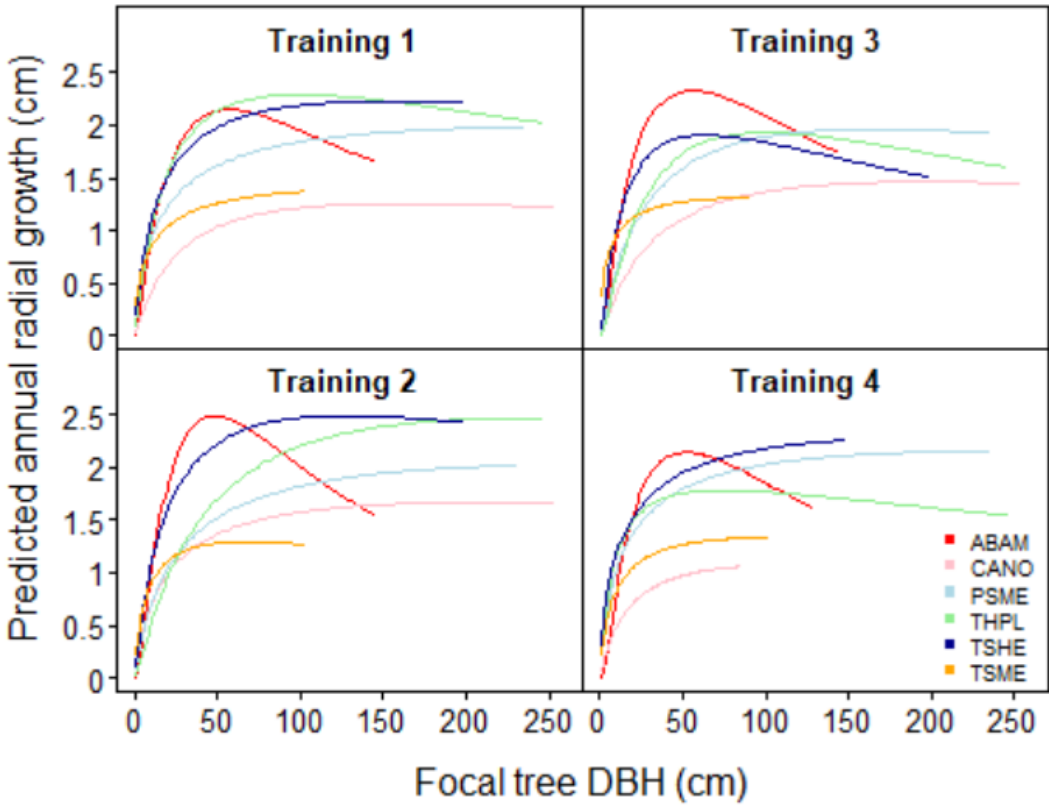

Figure S1.1. Fitted relationships between focal tree size and growth for the AIC likelihood models. Each line reflects the best AIC likelihood model for the indicated focal species by training set combination.

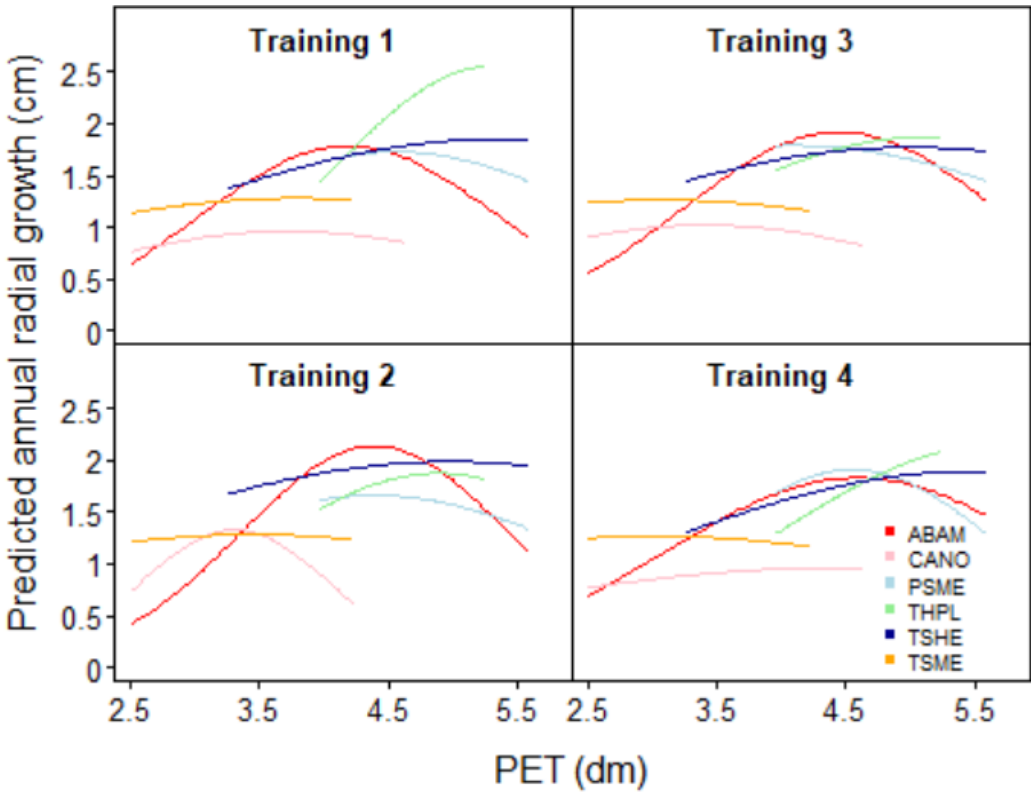

Figure S1.2. Fitted relationships between potential evapotranspiration (PET) and growth for the AIC likelihood models. Each line reflects the best AIC likelihood model for the indicated focal species by training set combination.

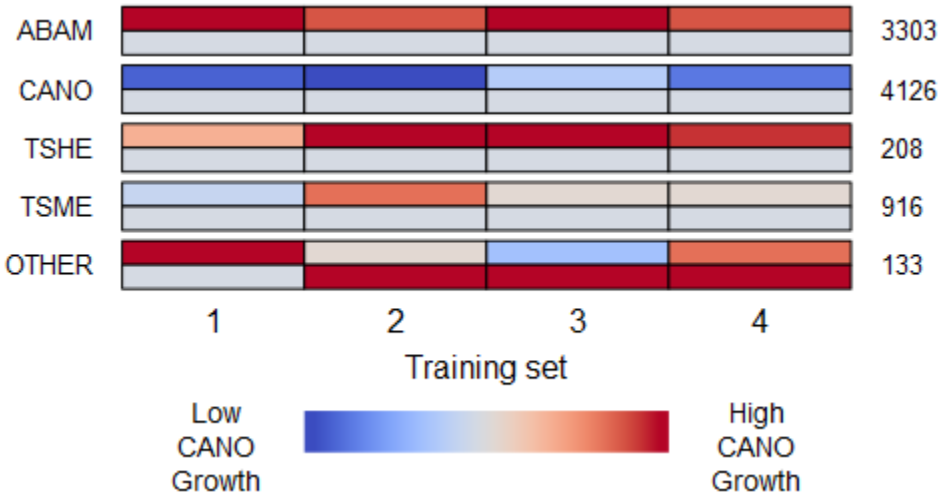

Figure S1.3. Which neighbor species are associated with highest/lowest growth of CANO focals? For each neighbor species there are two rows of colored bars. The top row of bars shows the likelihood model results and the bottom row shows the regularized regression model results. Each row of bars is divided into four sections, to show the results according to the models fit to each of the four training sets. The color of the bars indicates the growth rate of CANO in the presence of the neighbor species that row represents (see inset legend). The numbers on the right of the figure indicate the number of neighbors of each species averaged across training sets.

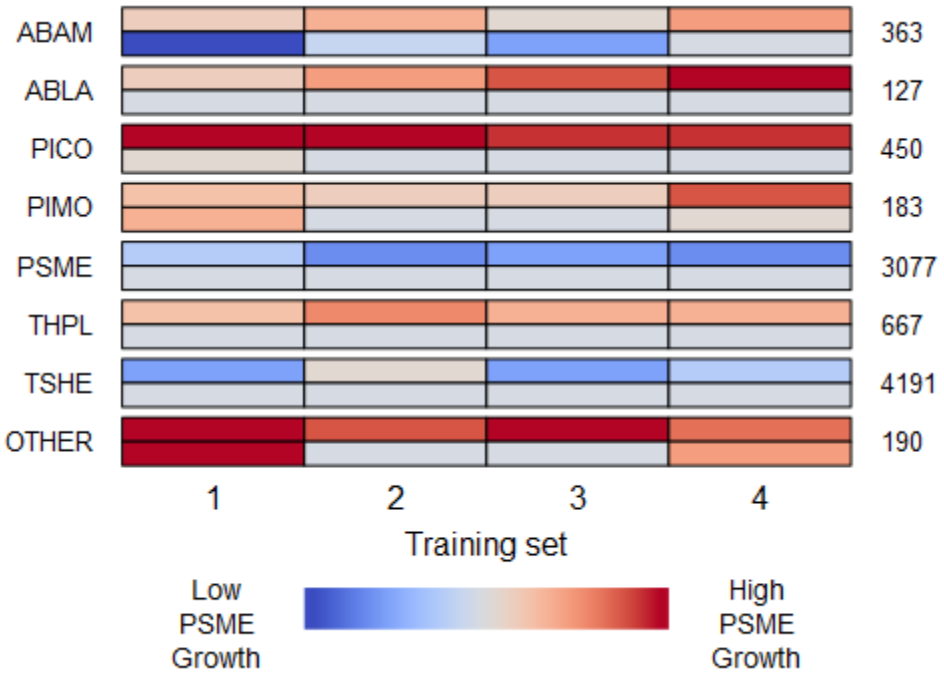

Figure S1.4. Which neighbor species are associated with highest/lowest growth of PSME focals? For each neighbor species there are two rows of colored bars. The top row of bars shows the likelihood model results and the bottom row shows the regularized regression model results. Each row of bars is divided into four sections, to show the results according to the models fit to each of the four training sets. The color of the bars indicates the growth rate of PSME in the presence of the neighbor species that row represents (see inset legend). The numbers on the right of the figure indicate the number of neighbors of each species averaged across training sets.

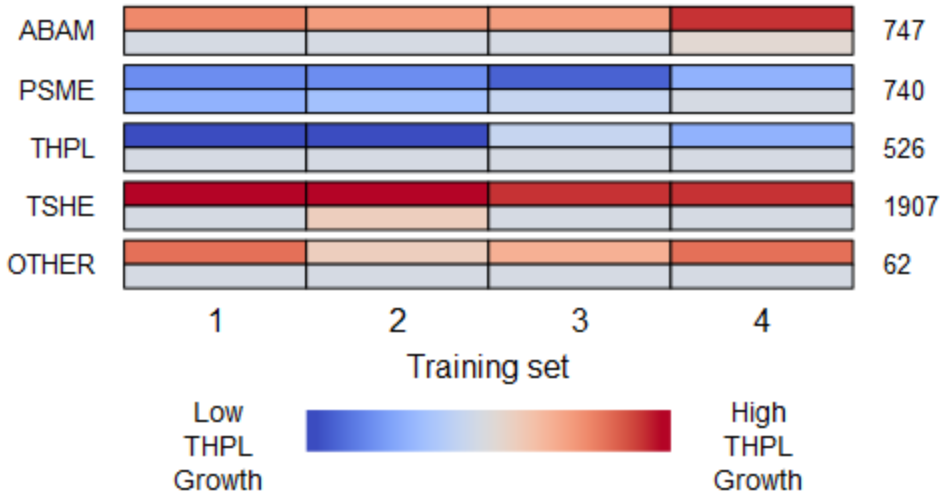

Figure S1.5. Which neighbor species are associated with highest/lowest growth of THPL focals? For each neighbor species there are two rows of colored bars. The top row of bars shows the likelihood model results and the bottom row shows the regularized regression model results. Each row of bars is divided into four sections, to show the results according to the models fit to each of the four training sets. The color of the bars indicates the growth rate of THPL in the presence of the neighbor species that row represents (see inset legend). The numbers on the right of the figure indicate the number of neighbors of each species averaged across training sets.

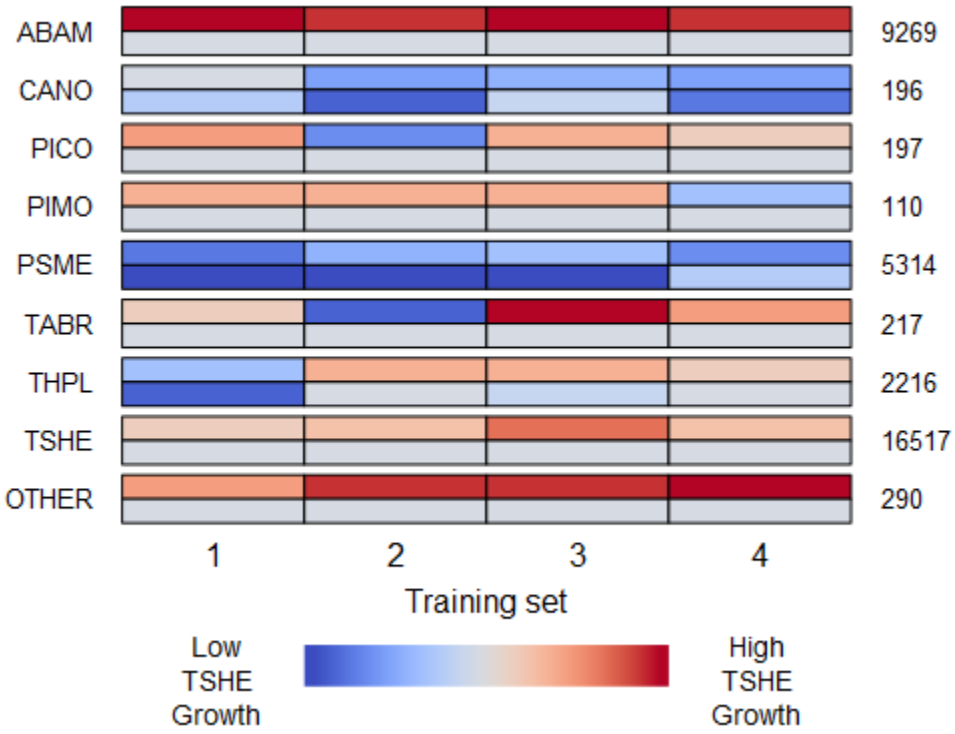

Figure S1.6. Which neighbor species are associated with highest/lowest growth of TSHE focals? For each neighbor species there are two rows of colored bars. The top row of bars shows the likelihood model results and the bottom row shows the regularized regression model results. Each row of bars is divided into four sections, to show the results according to the models fit to each of the four training sets. The color of the bars indicates the growth rate of TSHE in the presence of the neighbor species that row represents (see inset legend). The numbers on the right of the figure indicate the number of neighbors of each species averaged across training sets.

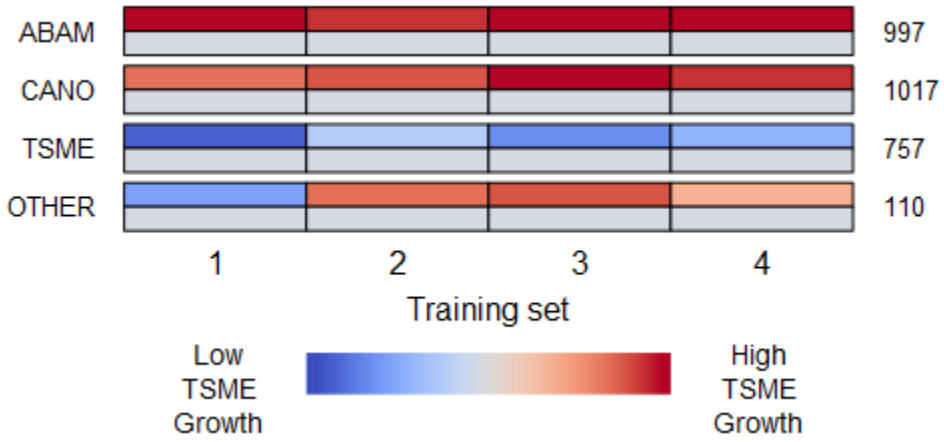

Figure S1.7. Which neighbor species are associated with highest/lowest growth of TSME focals? For each neighbor species there are two rows of colored bars. The top row of bars shows the likelihood model results and the bottom row shows the regularized regression model results. Each row of bars is divided into four sections, to show the results according to the models fit to each of the four training sets. The color of the bars indicates the growth rate of TSME in the presence of the neighbor species that row represents (see inset legend). The numbers on the right of the figure indicate the number of neighbors of each species averaged across training sets.

Table S1.1. Species included in neighborhood analyses as either focal trees or neighbors. In the main text and the supplement, all species are referred to by their four-letter code presented in this table. Some species were not modeled as focal trees but did appear as neighbors in the models for other species.

Species name	Code	Role in analyses
Abies amabilis	ABAM	Focal and neighbor
Abies lasiocarpa	ABLA	Neighbor only
Chamaecyparis nootkatensis	CANO	Focal and neighbor
Pinus contorta	PICO	Neighbor only
Pinus monticola	PIMO	Neighbor only
Pseudotsuga menziesii	PSME	Focal and neighbor
Taxus brevifolia	TABR	Neighbor only
Thuja plicata	THPL	Focal and neighbor
Tsuga heterophylla	TSHE	Focal and neighbor
Tsuga mertensiana	TSME	Focal and neighbor

Table S1.2. Summary of focal species' abundances, sizes, and densities across the sampled forest plots. The "Number of plots" column indicates the number of sampled plots in which each focal species occurred. The remaining columns summarize the abundance, size and density of each species, represented as the mean across the plots in which the species occurred, with standard deviation in parentheses. Standard deviations are large because the abundance of trees varied dramatically between plots as a result of their elevational distributions. For DBH values, a mean DBH was calculated for each species in each plot and values in this table reflect the mean and standard deviation of these within-plot means. The contribution of each tree to density was calculated using its area at breast height. Proportional densities represent the proportion of total tree density constituted by the indicated focal species.

Focal Species	Number of plots	Number of trees	Mean DBH (cm)	Density (m^2/ha)	Proportional density
ABAM	14	124 (161)	24.7 (12.7)	11.3 (11.0)	0.24 (0.23)
CANO	7	44 (58)	51.7 (40.0)	6.9 (7.5)	0.16 (0.17)
PSME	10	35 (40)	120.1 (66.9)	17.6 (13.7)	0.36 (0.21)
THPL	9	17 (15)	87.4 (58.1)	8.4 (6.2)	0.16 (0.10)
TSHE	13	107 (56)	35.7 (12.9)	15.0 (7.9)	0.32 (0.18)
TSME	4	32 (36)	46.2 (26.1)	7.3 (8.5)	0.18 (0.20)

Table S1.3. AIC Likelihood model selection results. For each focal species by training set combination, we fit four likelihood models differing in the structure of the crowding effect. Values in this table represent ΔAICc values for each of the four models, with the bolded '0's indicating the best model for each focal species by training set combination.

Focal species	Training set	No interactions	Equivalent interactions	Intraspecific vs. interspecific interactions	Species-specific interactions
ABAM	1	186.32	15.24	**0**	9.35
	2	164.8	27.34	**0**	0.11
	3	162.53	30.82	**0**	10.48
	4	172.36	30.95	8.89	**0**
CANO	1	36.85	24.72	**0**	1.54
	2	51.27	19.93	**0**	9.22
	3	36.99	19.22	**0**	2.03
	4	39.1	16.46	**0**	11.54
PSME	1	63.59	**0**	5.31	5.18
	2	58.33	0.71	**0**	5.57
	3	63.81	0.75	3.44	**0**
	4	47.68	2.29	**0**	7.88
THPL	1	35.14	8.7	10.17	**0**
	2	51.19	26.76	11.75	**0**
	3	26.86	7.07	1.5	**0**
	4	55.22	26.67	27.23	**0**
TSHE	1	184.94	112.62	93.65	**0**
	2	148.56	97.88	85.81	**0**
	3	152.34	78.72	66.52	**0**
	4	187.18	106.54	95.36	**0**
TSME	1	5.77	13.75	**0**	6.1
	2	14.63	20.83	**0**	5.9
	3	15.83	10.25	**0**	7.26
	4	6.86	13.7	**0**	3.43

Table S1.4. Fitted parameter values for AIC likelihood models of ABAM. Values shown reflect the final model resulting from AICc model selection for each training set. NA values indicate that the parameter was not included in the final model structure.

Parameter	Training 1	Training 2	Training 3	Training 4
X0	5.647184	4.782863	5.570202	5.209209
Xb	1.296231	1.130307	1.247898	1.200916
gmax	4.898073	5.053127	5.013657	4.564527
pet_a	4.199276	4.398411	4.455179	4.593256
pet_b	1.191779	1.04632	1.23811	1.504272
C	0.77787	0.423502	0.256277	0.361589
α	0.794951	0.727756	0.516272	0.35993
β	1.368613	0.967788	1.068217	0.996048
λ_{con}	0.10418	0.096067	0.251779	NA
λ_{het}	0.183867	0.162946	0.522093	NA
λ_{ABAM}	NA	NA	NA	0.134526
λ_{ABLA}	NA	NA	NA	NA
λ_{CANO}	NA	NA	NA	0.526799
λ_{OTHR}	NA	NA	NA	0.096021
λ_{PICO}	NA	NA	NA	NA
λ_{PIMO}	NA	NA	NA	NA
λ_{PSME}	NA	NA	NA	0.988057
λ_{TABR}	NA	NA	NA	0.062519
λ_{THPL}	NA	NA	NA	0.908878
λ_{TSHE}	NA	NA	NA	0.349803
λ_{TSME}	NA	NA	NA	0.696702
σ	0.891298	0.874871	0.88144	0.861362

Table S1.5. Fitted parameter values for AIC likelihood models of CANO. Values shown reflect the final model resulting from AICc model selection for each training set. NA values indicate that the parameter was not included in the final model structure.

Parameter	Training 1	Training 2	Training 3	Training 4
X0	16.31197	22.46111	19.58499	12.44248
Xb	1.896297	2.36642	1.848483	2.02088
gmax	3.351031	4.942072	3.581512	3.205381
pet_a	3.727071	3.308012	3.413579	4.453424
pet_b	1.830967	0.747256	1.854592	2.930325
C	0.091332	0.190015	0.038554	0.015167
α	1.830106	0.594843	1.198453	1.521665
β	0.562342	0.748868	0.527449	0.125047
λ_{con}	0.13695	0.565522	0.672332	0.575118
λ_{het}	0.000868	0.008808	0.02458	0.007078
λ_{ABAM}	NA	NA	NA	NA
λ_{ABLA}	NA	NA	NA	NA
λ_{CANO}	NA	NA	NA	NA
λ_{OTHR}	NA	NA	NA	NA
λ_{PICO}	NA	NA	NA	NA
λ_{PIMO}	NA	NA	NA	NA
λ_{PSME}	NA	NA	NA	NA
λ_{TABR}	NA	NA	NA	NA
λ_{THPL}	NA	NA	NA	NA
λ_{TSHE}	NA	NA	NA	NA
λ_{TSME}	NA	NA	NA	NA
σ	0.699108	0.620747	0.744168	0.697623

Table S1.6. Fitted parameter values for AIC likelihood models of PSME. Values shown reflect the final model resulting from AICc model selection for each training set. NA values indicate that the parameter was not included in the final model structure.

Parameter	Training 1	Training 2	Training 3	Training 4
X0	29.05303	29.51337	16.63528	25.60949
Xb	2.735194	2.331625	1.860527	2.681528
gmax	5.731953	6.11598	4.647375	5.539137
pet_a	4.557838	4.389105	4.093686	4.548516
pet_b	1.721984	1.827088	2.301967	1.177249
C	0.017006	0.031851	0.034251	0.091242
α	0.8687	0.695544	0.700755	0.961487
β	0.291482	0.278451	0.437992	0.133593
λ_{con}	NA	0.787429	NA	0.107496
λ_{het}	NA	0.624314	NA	0.096926
λ_{ABAM}	NA	NA	0.458198	NA
λ_{ABLA}	NA	NA	0.122873	NA
λ_{CANO}	NA	NA	NA	NA
λ_{OTHR}	NA	NA	0.044155	NA
λ_{PICO}	NA	NA	0.050529	NA
λ_{PIMO}	NA	NA	0.419464	NA
λ_{PSME}	NA	NA	0.756168	NA
λ_{TABR}	NA	NA	NA	NA
λ_{THPL}	NA	NA	0.335739	NA
λ_{TSHE}	NA	NA	0.78104	NA
λ_{TSME}	NA	NA	NA	NA
σ	0.889206	0.84796	0.869906	0.984844

Table S1.7. Fitted parameter values for AIC likelihood models of THPL. Values shown reflect the final model resulting from AICc model selection for each training set. NA values indicate that the parameter was not included in the final model structure.

Parameter	Training 1	Training 2	Training 3	Training 4
X0	9.992893	23.44913	9.988668	7.549333
Xb	1.818457	1.791436	1.475075	2.207433
gmax	6.63176	7.102305	6.480974	6.661364
pet_a	5.299065	4.885047	5.117365	5.673229
pet_b	1.242201	1.432689	1.896622	1.70106
C	0.066707	0.022204	0.014774	0.084291
α	0.671227	1.539787	1.259407	0.98542
β	0.406334	0.629883	0.102811	0.629119
λ_{con}	NA	NA	NA	NA
λ_{het}	NA	NA	NA	NA
λ_{ABAM}	0.210538	0.283883	0.267641	0.097394
λ_{ABLA}	NA	NA	NA	NA
λ_{CANO}	NA	NA	NA	NA
λ_{OTHR}	0.190251	0.435897	0.340124	0.178892
λ_{PICO}	NA	NA	NA	NA
λ_{PIMO}	NA	NA	NA	NA
λ_{PSME}	0.83859	0.828972	0.916601	0.724695
λ_{TABR}	NA	NA	NA	NA
λ_{THPL}	0.984868	0.993763	0.558268	0.740402
λ_{TSHE}	0.030303	0.015027	0.091588	0.050796
λ_{TSME}	NA	NA	NA	NA
σ	1.527451	1.52001	1.315754	1.224182

Table S1.8. Fitted parameter values for AIC likelihood models of TSHE. Values shown reflect the final model resulting from AICc model selection for each training set. NA values indicate that the parameter was not included in the final model structure.

Parameter	Training 1	Training 2	Training 3	Training 4
X0	15.25017	12.71974	6.366752	24.08188
Xb	2.301613	1.977892	1.662411	2.754767
gmax	4.517549	4.53113	3.445734	5.228501
pet_a	5.329706	4.980551	5.002725	5.470161
pet_b	2.685357	2.945408	2.743225	2.562391
C	0.122307	0.036658	0.176239	0.104756
α	0.869691	0.960747	0.863298	1.004421
β	1.125491	0.514917	1.11789	1.016064
λ_{con}	NA	NA	NA	NA
λ_{het}	NA	NA	NA	NA
λ_{ABAM}	0.018176	0.070921	0.014686	0.085824
λ_{ABLA}	NA	NA	NA	NA
λ_{CANO}	0.548442	0.784716	0.722251	0.793071
λ_{OTHR}	0.282171	0.066991	0.066905	0.021081
λ_{PICO}	0.257224	0.800655	0.314535	0.435677
λ_{PIMO}	0.306511	0.311458	0.333223	0.658512
λ_{PSME}	0.861707	0.748909	0.674588	0.823778
λ_{TABR}	0.432688	0.921135	0.018978	0.250154
λ_{THPL}	0.667487	0.328821	0.317018	0.407334
λ_{TSHE}	0.422378	0.357736	0.165325	0.367659
λ_{TSME}	NA	NA	NA	NA
σ	1.268121	1.22943	1.251041	1.256734

Table S1.9. Fitted parameter values for AIC likelihood models of TSME. Values shown reflect the final model resulting from AICc model selection for each training set. NA values indicate that the parameter was not included in the final model structure.

Parameter	Training 1	Training 2	Training 3	Training 4
X0	20.1121	6.746346	10.33789	11.09961
Xb	2.974753	2.230119	2.933228	2.478097
gmax	2.426373	2.064813	2.2206	2.382811
pet_a	3.785202	3.445726	3.058415	3.013445
pet_b	2.664246	2.764165	2.799773	2.952137
C	0.071646	0.01912	0.033805	0.037185
α	2.500453	2.215656	2.137544	2.025811
β	1.294959	1.441635	1.638055	1.186808
λ_{con}	0.116883	0.991576	0.81256	0.548381
λ_{het}	0.011666	0.006159	0.108582	0.029573
λ_{ABAM}	NA	NA	NA	NA
λ_{ABLA}	NA	NA	NA	NA
λ_{CANO}	NA	NA	NA	NA
λ_{OTHR}	NA	NA	NA	NA
λ_{PICO}	NA	NA	NA	NA
λ_{PIMO}	NA	NA	NA	NA
λ_{PSME}	NA	NA	NA	NA
λ_{TABR}	NA	NA	NA	NA
λ_{THPL}	NA	NA	NA	NA
λ_{TSHE}	NA	NA	NA	NA
λ_{TSME}	NA	NA	NA	NA
σ	0.533293	0.582348	0.483584	0.507845

Table S1.10. Estimated coefficients for regularized regression models of ABAM. Positive coefficients indicate that focal tree growth was positively associated with the indicated variable. The "Neighbor species:" variables are binary and therefore a positive coefficient indicates that focal tree growth was greater when the Neighbor was of that species. The coefficients of some variables were reduced to zero through regularization, indicating that they do not have a strong effect on focal tree growth.

Independent Variable	Training 1	Training 2	Training 3	Training 4
Neighbor species: ABAM	0.001387	0.000975	0.0011	0.001032
Neighbor species: CANO	0	0	0	-3.69E-05
Neighbor species: OTHR	0	0	0	2.95E-05
Neighbor species: PSME	-0.0009	-0.00111	-0.00126	-0.00141
Neighbor species: TABR	0	0	0	0
Neighbor species: THPL	-0.00058	-5.80E-06	-0.00052	-0.00061
Neighbor species: TSHE	0	0	0	0
Neighbor species: TSME	-0.00032	-0.00026	-0.0004	-0.00023
Neighbor size	0.00039	0.000172	0.000418	0.000647
Neighbor proximity	0	0	5.48E-06	9.28E-05
Local density: ABAM	0	0	0	0
Local density: CANO	-0.00517	-0.00576	-0.00509	-0.00525
Local density: OTHR	-0.00086	-0.00037	-0.00085	0
Local density: PSME	-0.00083	-0.00033	-0.00028	-0.00111
Local density: TABR	-0.00419	-0.00755	-0.00617	-0.00634
Local density: THPL	0	0	-0.00095	-0.00114
Local density: TSHE	0.008367	0.008489	0.008838	0.009301
Local density: TSME	-0.00166	-0.00168	-0.00124	-0.00147
Local density: All species	-0.00111	-0.00091	0	-0.00045
PET	0.004492	0.005472	0.006743	0.005741

Table S1.11. Estimated coefficients for regularized regression models of CANO. Positive coefficients indicate that focal tree growth was positively associated with the indicated variable. The "Neighbor species:" variables are binary and therefore a positive coefficient indicates that focal tree growth was greater when the Neighbor was of that species. The coefficients of some variables were reduced to zero through regularization, indicating that they do not have a strong effect on focal tree growth.

Independent Variable	Training 1	Training 2	Training 3	Training 4
Neighbor species: ABAM	0	0	0	0
Neighbor species: CANO	0	0	0	0
Neighbor species: OTHR	0	0.002186	0.002229	0.001159
Neighbor species: TSHE	0	0	0	0
Neighbor species: TSME	0	0	0	0
Neighbor size	0	0	0	0
Neighbor proximity	0	0	0	0
Local density: ABAM	-0.00063	0	0	0
Local density: CANO	-0.00819	-0.01013	-0.01119	-0.01049
Local density: OTHR	0	0	0.000631	0.001404
Local density: TSHE	0.001136	0.000435	0	0.001029
Local density: TSME	0.000331	0	0	0
Local density: All species	0	0	0	0
PET	0.00235	0.001789	0.002703	0.00343

Table S1.12. Estimated coefficients for regularized regression models of PSME. Positive coefficients indicate that focal tree growth was positively associated with the indicated variable. The "Neighbor species:" variables are binary and therefore a positive coefficient indicates that focal tree growth was greater when the Neighbor was of that species. The coefficients of some variables were reduced to zero through regularization, indicating that they do not have a strong effect on focal tree growth.

Independent Variable	Training 1	Training 2	Training 3	Training 4
Neighbor species: ABAM	-0.00038	-0.00014	-0.00024	0
Neighbor species: ABLA	0	0	0	0
Neighbor species: OTHR	0.000449	0	0	0.000211
Neighbor species: PICO	8.01E-05	0	0	0
Neighbor species: PIMO	0.000202	0	0	4.11E-05
Neighbor species: PSME	0	0	0	0
Neighbor species: THPL	0	0	0	0
Neighbor species: TSHE	-4.66E-05	0	0	0
Neighbor size	0	0	0	0
Neighbor proximity	0	0	0	0
Local density: ABAM	-0.00081	-0.00116	-0.00195	-0.00075
Local density: ABLA	-0.00081	0	0.00089	0
Local density: OTHR	0.001114	0.000693	0.000446	0.000541
Local density: PICO	0.012849	0.012966	0.010673	0.014151
Local density: PIMO	0.008998	0.008083	0.007458	0.004956
Local density: PSME	0	0	0	0
Local density: THPL	0.002596	0.002899	0.002923	0.002573
Local density: TSHE	-0.00028	-0.00082	-0.00134	-0.00164
Local density: All species	-0.00551	-0.00497	-0.00553	-0.00419
PET	0.002586	0.002183	0.001461	0.000514

Table S1.13. Estimated coefficients for regularized regression models of THPL. Positive coefficients indicate that focal tree growth was positively associated with the indicated variable. The "Neighbor species:" variables are binary and therefore a positive coefficient indicates that focal tree growth was greater when the Neighbor was of that species. The coefficients of some variables were reduced to zero through regularization, indicating that they do not have a strong effect on focal tree growth.

Independent Variable	Training 1	Training 2	Training 3	Training 4
Neighbor species: ABAM	0	0	0	0.000117
Neighbor species: OTHR	0	0	0	0
Neighbor species: PSME	-0.00076	-0.00027	-0.00038	0
Neighbor species: THPL	0	0	0	0
Neighbor species: TSHE	0	0.000113	0	0
Neighbor size	0	0	0	0
Neighbor proximity	0	0	0	0
Local density: ABAM	0	0	0.001481	0.000169
Local density: OTHR	0.001395	0.003349	0.003272	0.003685
Local density: PSME	0	-0.00565	-7.14E-05	0
Local density: THPL	0	-0.00605	0	0
Local density: TSHE	0.000217	0.002504	-0.00078	0
Local density: All species	-0.00844	0	-0.00694	-0.00554
PET	0.004381	0.001116	0	0.00059

Table S1.14. Estimated coefficients for regularized regression models of TSHE. Positive coefficients indicate that focal tree growth was positively associated with the indicated variable. The "Neighbor species:" variables are binary and therefore a positive coefficient indicates that focal tree growth was greater when the Neighbor was of that species. The coefficients of some variables were reduced to zero through regularization, indicating that they do not have a strong effect on focal tree growth.

Independent Variable	Training 1	Training 2	Training 3	Training 4
Neighbor species: ABAM	0	0	0	0
Neighbor species: CANO	-0.00029	-0.00061	-0.00044	-0.00087
Neighbor species: OTHR	0	0	0	0
Neighbor species: PICO	0	0	0	0
Neighbor species: PIMO	0	0	0	0
Neighbor species: PSME	-0.00084	-0.00105	-0.00102	-0.0003
Neighbor species: TABR	0	0	0	0
Neighbor species: THPL	-0.00056	-0.00025	-0.00043	-6.91E-05
Neighbor species: TSHE	0	0	0	0
Neighbor size	0	0	0	0
Neighbor proximity	0	0	0	0
Local density: ABAM	0.016833	0.014725	0.016444	0.012713
Local density: CANO	0	0	0	0
Local density: OTHR	0	-0.00045	0	0
Local density: PICO	0.00413	0.000991	0.000203	0
Local density: PIMO	-0.00343	0	-0.00013	0
Local density: PSME	-0.01013	-0.00818	-0.00949	-0.01179
Local density: TABR	0.000579	0.001252	0.001889	0.000582
Local density: THPL	0	0	0	0
Local density: TSHE	-0.00522	-0.00561	-0.00453	-0.00573
Local density: All species	-0.00495	-0.00564	-0.00563	-0.0051
PET	0.011216	0.009063	0.011029	0.01117

Table S1.15. Estimated coefficients for regularized regression models of TSME. Positive coefficients indicate that focal tree growth was positively associated with the indicated variable. The "Neighbor species:" variables are binary and therefore a positive coefficient indicates that focal tree growth was greater when the Neighbor was of that species. The coefficients of some variables were reduced to zero through regularization, indicating that they do not have a strong effect on focal tree growth.

Independent Variable	Training 1	Training 2	Training 3	Training 4
Neighbor species: ABAM	0	0	0	0
Neighbor species: CANO	0	-0.00031	-0.00038	-0.00022
Neighbor species: OTHR	0	0	0	0
Neighbor species: TSME	0	0	0	0
Neighbor size	0	0	0	0
Neighbor proximity	0	0	0	0
Local density: ABAM	0	0	0	0
Local density: CANO	0	-0.0023	-0.01193	-0.00152
Local density: OTHR	0.003089	0.001238	0	0.001754
Local density: TSME	-0.00799	-0.00915	-0.01324	-0.00752
Local density: All species	-0.00537	-0.00289	-0.00052	-0.00585
PET	0	0	0	0

Table S1.16. Is focal growth higher in the presence of conspecific or heterospecific neighbors? For the AIC likelihood model, values are the optimized heterospecific interaction coefficient (λ_{het}) minus the optimized conspecific interaction coefficient (λ_{con}), both taken from the best fitting conspecific vs. heterospecific interactions model. For the regularized regression model, the direction of the effect is shown (see main text Table 1 for details). For both model types, positive values indicate that focal growth was higher in the presence of conspecific neighbors. For each regularized regression model, the sign of any effect indicated was observed in all of the 100 models that were run. NA values for the regularized regression indicate that the model did not focal trees to grow substantially faster in the presence of conspecific or heterospecific neighbors. Note that this table provides the numerical output underlying main text Table 1.2.

Focal Species	AIC Likelihood				Regularized Regression			
	1	**2**	**3**	**4**	**1**	**2**	**3**	**4**
ABAM	0.08	0.07	0.27	0.47	+	+	+	+
CANO	-0.14	-0.56	-0.65	-0.57	-	-	-	-
PSME	0.07	-0.16	-0.09	-0.01	NA	NA	NA	NA
THPL	-0.23	-0.25	0	-0.65	NA	-	NA	NA
TSHE	-0.32	-0.4	-0.11	-0.44	-	-	-	-
TSME	-0.11	-0.99	-0.7	-0.52	-	-	-	-

Table S1.17. Amount of variance in the training and test datasets explained by the AIC likelihood (L-AIC), CV likelihood (L-CV) and regularized regression (RR) models. Sample size reflects the number of focal trees in the training set. Values in the six rightmost columns are coefficients of determination, which are the proportion of variance around the mean tree growth value explained by the model. The maximum possible value for a coefficient of determination is 1 (all variance explained) but negative values can exist when a model is applied to unseen test data if there is more unexplained variation around model predictions than exists around the mean growth value in the test data.

Focal species	Training set	Sample size	Training fit			Test fit		
			L-AIC	L-CV	RR	L-AIC	L-CV	RR
ABAM	1	1297	0.38	0.38	0.15	0.4	0.4	0.21
	2	1307	0.41	0.41	0.19	0.32	0.32	0.12
	3	1310	0.39	0.4	0.18	0.38	0.36	0.13
	4	1303	0.39	0.39	0.16	0.39	0.39	0.18
CANO	1	230	0.28	0.12	0.18	0.22	-0.06	0.25
	2	227	0.34	0.34	0.21	-0.01	-0.01	0.24
	3	233	0.25	0.25	0.25	0.28	0.28	0.16
	4	234	0.22	0.22	0.26	0.24	0.24	0.08
PSME	1	267	0.34	0.34	0.63	0.31	0.3	0.58
	2	264	0.35	0.37	0.64	0.29	0.3	0.61
	3	259	0.37	0.11	0.65	0.4	0.2	0.57
	4	251	0.34	0.34	0.59	0.31	0.31	0.7
THPL	1	112	0.51	0.51	0.17	0.5	0.5	0.01
	2	117	0.57	0.57	0.19	0.17	0.17	0.06
	3	120	0.45	0.45	0.17	0.55	0.55	-0.02
	4	119	0.6	0.6	0.15	0.28	0.28	0.1
TSHE	1	1043	0.3	0.3	0.27	0.25	0.25	0.18
	2	1045	0.27	0.27	0.23	0.34	0.34	0.27
	3	1043	0.28	0.21	0.24	0.26	0.18	0.25
	4	1054	0.31	0.31	0.24	0.23	0.23	0.23
TSME	1	97	0.24	0.24	0.24	0.29	0.26	-0.04
	2	96	0.28	0.28	0.15	-0.03	-0.03	0.34
	3	89	0.3	0.04	0.26	0.07	0.03	-0.1
	4	96	0.24	0.07	0.21	0.17	-0.16	0.1

Table S1.18. CV Likelihood model selection results. For each focal species by training set combination, four likelihood models differing in the structure of the crowding effect were fitted. Values in this table represent mean square error values for each of the four models, averaged across the 10 folds used in cross-validation. Bolded values indicate the best model (lowest mean square error) for each focal species by training set combination.

Focal species	Training set	No interactions	Equivalent interactions	Intraspecific vs. interspecific interactions	Species-specific interactions
ABAM	1	0.82	0.76	**0.74**	0.78
	2	0.72	0.81	**0.65**	0.67
	3	0.76	0.71	0.70	**0.65**
	4	0.90	0.88	0.87	**0.84**
CANO	1	**0.34**	0.46	0.44	0.40
	2	0.37	0.20	**0.19**	0.21
	3	0.51	0.39	**0.32**	0.36
	4	0.64	0.62	**0.41**	0.48
PSME	1	1.30	1.08	**0.94**	0.96
	2	0.90	0.58	0.62	**0.53**
	3	**1.02**	1.05	1.07	1.10
	4	0.96	0.84	**0.75**	0.76
THPL	1	2.99	2.24	1.96	**1.40**
	2	2.91	1.70	1.60	**1.22**
	3	1.35	1.03	0.96	**0.87**
	4	3.46	4.27	2.70	**1.21**
TSHE	1	2.12	2.03	1.98	**1.89**
	2	2.01	1.87	1.87	**1.65**
	3	1.72	**1.59**	1.59	1.64
	4	2.24	2.13	2.06	**1.91**
TSME	1	0.11	0.10	0.09	**0.08**
	2	0.29	0.27	**0.23**	0.25
	3	**0.25**	0.35	0.31	0.34
	4	**0.68**	0.73	0.80	0.88

Chapter 2. ForestPlotR: an R package for the exploration and analysis of stem-mapped forest stand data

For submission to Ecography: Stuart I. Graham, Ariel Rokem, Janneke Hille Ris Lambers

Abstract

Stem-mapped forest stands offer important opportunities for investigating the fine-scale spatial processes occurring in forest ecosystems. These stands are areas of the forest where the precise locations and repeated size measurements of each tree are recorded, thereby enabling the calculation of spatially-explicit metrics of tree growth rates and of the tree community. The most common use of these rich datasets is to investigate the drivers of variation in tree growth by modeling tree growth rate as a function of the aforementioned neighborhood metrics, but these metrics could also serve as important covariates of many other spatially variable patterns that influence trees or are influenced by trees, including herbivory and soil microbial community composition. Widespread use of stem-mapped forest stand datasets is currently hampered by the lack of standardized, efficient and easy-to-use tools for computing with these data. We present the ForestPlotR package that facilitates the munging, exploration, visualization, and analysis of stem-mapped forest stands. By providing flexible, user-friendly functions to calculate neighborhood metrics and implementation of a recently-developed rapid-fitting tree growth model, ForestPlotR broadens the accessibility of stem-mapped forest stand data. We demonstrate the functionality of ForestPlotR by using it to investigate how the species identity of neighboring trees influences the growth rates of three common tree species in Mount Rainier National Park, WA, USA. ForestPlotR is designed to facilitate researchers to incorporate spatially-explicit descriptions of tree communities in their studies and we expect this increased diversity of contributors to spawn exciting new ways of using stem-mapped forest stand data.

2.1. Introduction

Forest ecosystems represent a critical buffer against anthropogenic climate change and we therefore must develop a deep understanding of their dynamics. An important tool in this endeavor is the stem-mapped forest stand; an area of forest where the size, species identity, and precise location of each tree meeting a minimum size threshold are recorded (Condit, 1995). Once established, these stands are revisited periodically to document mortality, re-measure trees, and record any additional trees that newly meet the minimum size threshold. In particular, stem-mapped forest stands ≥ 1 ha in area (e.g., those of the Forest Global Earth Observatory (ForestGEO; Davies et al. 2021) and Pacific Northwest Permanent Sample Plot networks (Franklin et al. 2021)) offer important opportunities for investigating the fine-scale spatial processes that drive forest dynamics (Lutz, 2015). This is because they allow us to quantify the local tree community (hereafter "neighborhood" for consistency with published literature) surrounding any particular tree in a spatially-explicit way. Neighborhood metrics have been shown to be highly informative covariates of spatial variation in tree growth and survival (e.g., Uriarte et al. 2004, Wiegand et al. 2017), and have facilitated research on climate change (e.g., Buechling et al. 2017), competitive interactions (e.g., Fortunel et al. 2016) and pest-induced tree mortality (e.g., Buonanduci et al. 2020).

To extract these insights from stem-mapped forest stand datasets, a considerable amount of programming skill and computational power is required. The first task is to clean and format the data for analysis. This is a non-trivial data management problem because errors are common in field-collected data and each group of forest stands tends to have its own data structure. Next, the data must be explored to identify the important neighborhood metrics; an iterative process given that the appropriate neighborhood size is generally not known a priori, as it depends on local environment, the spatial process being explored (e.g. competition vs. soil microbiomes), species identity, and many other factors. In addition, calculating neighborhood metrics requires identifying all trees within the neighborhood of each tree, typically generating datasets with hundreds of thousands of rows; a daunting workflow for an inexperienced programmer. The final step is to build models for hypothesis testing that incorporate neighborhood

metrics, and many of these models require high-performance computing resources (Fortunel et al., 2018; Kunstler et al., 2016; Uriarte et al., 2004), to which not all researchers have access. Software tools that can alleviate these challenges would present an opportunity to increase the abundance and diversity of projects using these rich datasets.

Software tools for interacting with stem-mapped forest stand data are available, but their design restricts their use to specific dataset structures and analytical approaches. For example, the fgeo suite of R packages (Lepore et al. 2019) facilitates analyses of data from the ForestGEO network of stem-mapped forest stands (Davies et al., 2021). However, fgeo does not contain any functions for calculating neighborhood metrics, and its powerful habitat-species association analyses require individual tree-level elevation data, which are currently not available for many stands outside the ForestGEO network. The scope of the fgeo package is understandably limited in order to encourage researchers to explore all aspects of this diverse data source, but we believe there is a great deal of insight to be gained from tools that use only the data types common to all stem-mapped forest stands: the size, species identity, and precise locations of trees. Also of note is the rFIA package (Stanke et al., 2020), which supports the retrieval and analysis of data from the USFS Forest Inventory and Analysis (FIA) stands. However, the FIA stands, and all other circular stands we are aware of, are too small to encapsulate entire neighborhoods in most forest ecosystems (Lutz, 2015), but designed instead to cover a broad distribution and thereby permit analysis of tree performance patterns across large environmental gradients.

To address this gap in the current technical tool landscape, we present the ForestPlotR package, which facilitates the exploration, visualization, and analysis of rectangular stem-mapped forest stand data. ForestPlotR is specifically designed to identify and calculate important neighborhood metrics for use in further analyses and therefore requires data from forest stands large enough to capture complete neighborhoods (typically, neighborhood radius $\geq$ 10 m; Uriarte et al. 2004, Fortunel et al. 2016). ForestPlotR is compatible with all rectangular stem-mapped forest stand datasets because it requires only the common data types of the size, species identity, and position of trees, yet can also accommodate as

many additional variables as desired. It provides functions for calculating common neighborhood metrics, and creates clean visualizations for data exploration. In addition, ForestPlotR implements a new, efficient neighborhood model of tree growth that can be used to select an appropriate neighborhood size, guide the development of a more mechanistic model of tree growth, and test ecological hypotheses (Graham et al., 2021).

2.2. Package structure and functionality

The functions included in the ForestPlotR package are divided into five categories based on the analytical task they are designed to assist with (Figure 2.1). All users should begin with the functions in the Data Formatting category, which verify the user-provided data are in a format that other ForestPlotR functions can handle. Functions of the *Describing Neighborhoods* category create neighborhoods of a user-defined size and calculate corresponding neighborhood metrics. There are also sets of functions for *Calculating Growth Rates* and *Visualizing Mapped Stands*. The *Modeling* category contains two functions; one that fits a neighborhood model of tree growth and another that helps the user determine a suitable neighborhood size to use in their analyses.

In this section, we describe functions included in the package under each of these categories, and list some of their potential uses. Figure 2.1 provides guidance on how the function categories can be combined. ForestPlotR also includes a series of built-in stem-mapped forest stand datasets from Mount Rainier National Park (Franklin et al. 2021) that can be used to test its functions (see "Examples" in each of the function help files).

2.2.1. *Data Formatting*

ForestPlotR requires stem-mapped forest stand data to be provided in two separate data frames; one containing tree mapping data (i.e. spatial coordinates), and the other containing tree measurement data. The decision to require separation of these two data types was made to avoid redundancy, as most trees will have multiple records of measurement data (one for each stand census) but only one record of

mapping data. It follows that each tree should populate only a single row in the mapping data frame, but will populate *n* rows in the tree measurement data frame, where *n* is the number of censuses in which the tree was measured. For some stem-mapped forest stands, there will be a separate file of tree measurement data for each census, but these can easily be combined into a single file using dplyr::bind_rows.

The functions *mapping_check* and *tree_check* are provided to verify that mapping and tree measurement data frames are in a format compatible with ForestPlotR functions. Both functions return a list of two elements: the first element is a data frame of the ID codes of all the trees that have a potential data issue and a short description of the issue (see function documentation for list of issues); the second element is a data frame summarizing the number and percentage of tree ID codes that had at least one issue, and that had each of the specific issues. These functions do not remove any trees with data issues because it is likely that many of the data issues can be resolved through reference to field notes, etc. It is the user's responsibility to remove any trees with issues that cannot be resolved before continuing analysis.

In the interest of flexibility, the mapping and tree measurement data frames can each contain as many columns as desired (and these columns are retained in the output where possible; see function documentation for details), but the *mapping_check* and *tree_check* functions will stop and return warning messages if any required columns are missing. The mapping data frame must contain the columns: tree_id (code uniquely identifying each tree), stand_id (name of stand in which the tree is located), species (species identity of the tree as a string of any length containing no spaces), x_coord, and y_coord (x and y coordinates of the tree, in meters from stand origin). The tree measurement data frame must contain the columns: tree_id, stand_id, species, year (year in which measurement was taken – if measurement date is recorded at a finer temporal resolution, please use functions from the lubridate package to extract year information), and dbh (diameter at breast height (hereafter, DBH) in cm). All ForestPlotR functions, except those that create plots of a single stand (i.e., *contour_plot* and *stand_map*), accept datasets containing information from multiple stands as input.

2.2.2. *Describing Neighborhoods*

The first step in calculating neighborhood metrics is to identify the set of trees in each neighborhood. The *neighborhoods* function achieves this using a mapping and a tree census data frame and a specified neighborhood radius (for assistance in selecting an appropriate neighborhood radius, see the *select_nbhd_size* function described in section 2.5 Modeling tree growth). It returns a new data frame where each neighborhood ID appears on multiple rows, with each row containing information on one of the trees in that neighborhood (i.e., growing within 'radius' of the neighborhood center). By default, a neighborhood is constructed for each individual tree in the mapping data frame (and neighborhood IDs are equal to focal tree IDs), but the function can instead construct neighborhoods centered on specific locations within the stand if a data frame of these locations is provided under the argument 'coords'. This is useful if some other information has been sampled at a specific location in the stand (e.g., seed trap, soil microbial community), and the goal is to model it as a function of the local tree community. In addition, if the provided mapping dataset contains data from multiple forest stands, the user can specify for which stands the neighborhoods should be calculated. The data frame output by *neighborhoods* serves as the input for a number of other functions in ForestPlotR, including *neighborhood_summary* and *site_by_species*.

The *neighborhood_summary* function is applied to the data frame output by *neighborhoods*, and calculates neighborhood metrics for the neighborhood of each tree or user-provided set of coordinates. Specifically, it calculates the species richness, overall tree density, and the density of each tree species in each neighborhood. The default density measure is in units of m^2 basal area per hectare, but the species-specific densities can also be converted to the proportions of total basal area (summed across trees of all species) that they constitute. Finally, this function also allows densities to instead be calculated with the angular method (Rouvinen & Kuuluvainen, 1997):

$$species\ A\ density = \sum_{i=1}^{N} \arctan\left(\frac{dbh_i}{distance_i}\right), \tag{2.1}$$

where N is the number of neighbors of species A in the neighborhood, dbh_i is the diameter at breast height of neighbor i, and $distance_i$ is the distance of the neighbor i from the neighborhood center. The angular method may be most appropriate when the differential effects of neighbor tree species depend on how close those neighbor trees are to the focal (e.g., if differences are driven by competition for soil moisture or nutrients; Contreras et al. 2011). The calculated densities can be used as model covariates to estimate the effects of specific neighbor species on a metric of interest.

ForestPlotR also contains some additional functions that can help the user acquire further neighborhood metrics. The *site_by_species* function takes the output of *neighborhoods* as input and creates a site-by-species matrix (presence/absence or abundance-weighted) where each site is a neighborhood. The resulting matrix could be used in conjunction with other datasets to calculate either phylogenetic or functional diversity (see R packages picante (Kembel et al., 2010) and FD (Laliberté and Legendre 2010; Laliberté et al. 2014)). The *tree_utm* function takes a mapping data frame and information on the orientation and location of stands (latitude and longitude of stand origins; see function documentation for details) and outputs Universal Transverse Mercator (UTM) coordinates for each individual tree in the mapping data frame. The user could then connect their stem-mapped forest stand dataset with remote-sensing data and extract, for each individual tree, detailed topographical, crown-structure, or normalized difference vegetation index (NDVI) data.

2.2.3. *Calculating growth rates*

The repeated censuses of stem-mapped forest stands enable the calculation of growth rate for each individual tree. ForestPlotR provides two functions that implement these calculations using a tree measurement data frame. The *growth_summary* function extracts the earliest and most recent measurements of each tree and calculates an average annual growth rate across this period (both DBH change and basal area increment). It is not uncommon to obtain illogical negative annual growth rates due to measurement error of tree DBH in stem-mapped forest stands, and so the function returns a warning

stating the number of trees that exhibited a negative growth rate. The output also contains, for each tree that had a non-negative annual growth rate, a size-corrected average annual growth rate, calculated as:

$$size\ corrected\ annual\ growth\ rate = \sqrt{\frac{average\ annual\ DBH\ change}{first\ DBH\ measurement}} \quad (2.2)$$

This transformed growth rate variable partially accounts for the non-linear relationship between tree size and growth rate and tends to more closely follow a normal distribution than raw growth rates, thereby enabling linear modeling of tree growth (Graham et al., 2021). ForestPlotR does not currently provide a function for calculating size-corrected basal area increments because our growth model has not been validated with such data, but the user could make such a calculation by adapting the underlying code for the *growth_summary* function. The *detailed_growth* function works similarly to *growth_summary* but instead returns separate growth rates for each tree between each pair of consecutive stand censuses. The greater temporal resolution in the growth rates calculated by the *detailed_growth* function enables exploration of how tree growth rates have changed over time, perhaps in relation to climatic events.

2.2.4. *Visualizing mapped stands*

ForestPlotR contains several functions for visualizing stem-mapped forest stands, which facilitate exploratory data analysis. The *size_dist* function uses a tree measurement data frame to graph the size-class distribution of one or more mapped stands. These distributions are frequently used in forest ecology and management to understand forest stand structure dynamics (i.e., stand succession). The ability to construct these size-class distributions for multiple stands simultaneously is valuable because *size_dist* outputs a multi-panel plot that facilitates comparison.

The *contour_plot* function creates an interactive contour map of the stand based on the value of any quantitative variable measured at multiple locations in the stand. For example, the *neighborhoods* and *neighborhood_summary* functions can be combined to calculate tree density at many locations within a given mapped stand, and then *contour_plot* can be used to generate a map of spatial variation in tree

density across the stand. The resulting map could be used to identify suitable sampling locations for a study investigating effects of canopy density. This function also offers an option for handling edge effects, which may be important if the contours are based on neighborhood metrics that cannot be calculated accurately at stand edges (e.g. tree density). The *contour_plot* function could also be used to generate contour plots of edaphic variables measured on a grid in the stand, which could be used to estimate edaphic conditions for each tree (or any sampling location in the stand) at high spatial resolution.

To facilitate repeat censuses of stem-mapped forest stands or the identification of suitable sample sites, ForestPlotR can also produce stem maps of stem-mapped forest stands. The *stand_map* function takes as input a mapping and a tree census data frame, each representing a single mapped stand, and returns a stem map of either the entire stand or a specified subsection of it, with individual trees represented by points with size proportional to their most recent DBH measurement.

2.2.5. *Modeling tree growth*

A major research opportunity afforded by stem-mapped forest stands is the ability to model tree growth as a function of the spatial structure of the surrounding neighborhood of trees. Tree growth models have provided insights on the processes underlying variation in tree growth, but common modeling techniques require a substantial amount of programming and access to high-performance computing resources. To encourage research in this area, the ForestPlotR function *growth_model* implements a new regularized regression model of tree growth that can be run in minutes on a personal laptop (Graham et al., 2021). Briefly, the efficiency of this model results from its linear structure, which can capture the inherently non-linear process of tree growth through a transformation of the dependent growth variable that accounts for the major non-linear relationship; that between focal tree size and growth rate.

The efficiency of the *growth_model* function leads to a number of potential uses. First, the resulting model can be used to test hypotheses regarding the processes underlying variation in tree growth (Graham et al., 2021). Second, the model can be run multiple times using different neighborhood sizes and plotting

mean square error of the resulting models to select a neighborhood size for the final model; a process that takes considerable time when using traditional modeling methods. Third, the regularized regression format of the implemented model means that it can indicate the most influential of several correlated independent variables. This means that *growth_model* could be used to select the most important climate variables to include in a more complex non-linear tree growth model (e.g. (Uriarte et al., 2004) and thereby streamline the model building process (see Graham et al. 2021 for details).

To use the *growth_model* function, a single data matrix containing both neighborhood and growth data needs to be constructed. This is achieved by joining the outputs of the *neighborhoods*, *neighborhood_summary* and *growth_summary* functions. There is also an option to handle rare neighbor tree species by grouping them together under the species identity of "RARE". Finally, if the user splits their data into a training and test set, *growth_model* can conduct test set validation of the final model, which is important if the model is intended for use as a predictive tool (Tredennick et al., 2021). The output contains information on model fit and estimated coefficients, and contains the model object that can be used to make predictions when passed to stats::predict. Further details on the use of *growth_model* are provided in the case study below and in the "Modeling tree growth" vignette of ForestPlotR.

An overlooked but important step in building a tree growth model is the selection of an appropriate neighborhood size. The neighborhood size that corresponds to neighborhood metrics of maximal explanatory power for tree growth will vary among ecosystems, forest successional stages, and study species (Lutz 2015), and it is therefore strongly advised that a sensitivity analysis be conducted for neighborhood size. The ForestPlotR function *select_nbhd_size* helps the user make an informed decision on the appropriate neighborhood size by fitting *growth_model* using different neighborhood sizes and comparing model fit. The case study presented in the next section provides more details on how to use this function and select a neighborhood size.

2.3. Case Study

To demonstrate the utility of ForestPlotR, we investigate how tree growth is influenced by the species identity of neighboring trees at Mount Rainier National Park, WA, USA. This analysis uses the built-in datasets of ForestPlotR, which originate from 15 rectangular 1 ha stem-mapped forest stands at Mt. Rainier that are managed by the Pacific Northwest Permanent Sample Plot Network (Franklin et al. 2021). A similar analysis has already been conducted on these data with ForestPlotR (Graham et al. 2021), but the analysis presented here differs by: focusing on a subset of the focal species (*Abies amabilis*, *Pseudotsuga menziesii*, *Tsuga heterophylla*), using the entire dataset for fitting each model (i.e., no data put aside for validation), and allowing neighborhood size to differ between focal species.

It is generally expected that tree growth rates are influenced by the species identity of neighboring trees, but the nature of these interactions can range from facilitative to competitive. Of particular interest to ecologists studying coexistence is how tree growth is influenced by conspecific neighbors, because reduced growth in the presence of conspecifics indicates negative density-dependent growth – an important mechanism for the maintenance of biodiversity. The biggest challenges in such analyses are: (1) calculating neighborhood metrics; (2) determining the appropriate neighborhood size; and (3) optimizing the typically complex models of tree growth. ForestPlotR provides simple functions that address each of these challenges. The following paragraphs describe how this analysis was conducted with ForestPlotR (for underlying code see: https://github.com/sgraham9319/ForestPlotR_case_study).

The first step is to select the appropriate neighborhood size, which can be achieved using the *select_nbhd_size* function. This function applies *growth_model* multiple times, with each run using neighborhood metrics calculated according to a different neighborhood size, and calculates the mean square error of each model. For this case study, we tried neighborhood sizes of 2, 4, 6, 8, 10, 12, 14, 16, 18, and 20 m in radius. To ensure that models using different neighborhood sizes maintain equal sample size, *select_nbhd_size* excludes all trees whose neighborhood overlaps the stand boundary according to the largest neighborhood size to be tested (i.e. in this case, all trees within 20 m of a stand boundary were excluded). We also specified that species-specific tree densities in the neighborhoods should be calculated

according to the proportional method of *neighborhood_summary* and included stand-level abiotic variables (i.e., the built-in dataset *stand_abiotic*) as covariates in the model.

This plots of mean square error versus neighborhood size output by *select_nbhd_size* are shown in (Figure 2.2A-C). It is expected that a larger neighborhood size will generally lead to a better fitting model because it contains more information on each neighborhood. There is a trade-off however, in that the larger the neighborhood size, the more focal trees need to be excluded from the analysis as a result of their neighborhood overlapping the stand boundary. It is desirable to model growth of as many focal trees as we can without dramatically increasing mean square error because this gives us greater confidence that our findings will be applicable beyond the sampled stands. Therefore, we recommend choosing the neighborhood size that, based on visual observation, lies at the "elbow" of these plots, which represents where further increases in neighborhood size lead to only modest improvements in model fit. Following this protocol, we selected neighborhood sizes of: *A. amabilis* = 12m, *P. menziesii* = 10m, *T. heterophylla* = 12m.

Next, we fit the final model for each species using the selected neighborhood size and including all trees whose neighborhood of this size does not overlap a stand boundary. Each model outputs two coefficients that are indicative of the effect of a particular neighbor species on the focal species; one for the neighbor species identity variable and another for the neighborhood metric of that neighbor species' density. To obtain a single value for the effect of each neighbor species on the growth of each focal species, we averaged these two related coefficients – these are the values shown in Figure 2.2D-F.

We find that *A. amabilis* grows more quickly in the presence of conspecifics, indicating positive density-dependent growth, but that *A. amabilis* responds even more positively to neighboring *T. heterophylla*. By contrast, *P. menziesii* and *T. heterophylla* do not exhibit either positive or negative density-dependent growth, but instead grow particularly quickly in the presence of *Pinus contorta* and *A. amabilis* neighbors respectively. These findings are mostly, but not entirely consistent with those reported previously

(Graham et al. 2021), but the present models are likely more accurate because they use a larger dataset and a neighborhood size tailored to the focal species.

2.4. Discussion

ForestPlotR equips forest ecology researchers of all skillsets with the tools they need to explore and analyze rectangular stem-mapped forest stand datasets, and thereby has the potential to spur exciting new ways of utilizing this rich data source. Widespread use of stem-mapped forest stand datasets is currently hampered by the high programming and computational demands involved, particularly in the quantitative description of neighborhoods, selecting an appropriate neighborhood size, and fitting neighborhood models of tree performance. ForestPlotR removes these barriers by providing flexible and user-friendly functions for describing neighborhoods, selecting neighborhood size, and implementing a rapid-fitting tree growth model (Graham et al. 2021) that can be used to investigate species interactions. Moreover, by requiring only the data types common to all stem-mapped forest stands (species identity, location, DBH measurements), ForestPlotR is compatible with all rectangular stem-mapped forest stand datasets. This means that stem-mapped forest stand datasets are now accessible to all researchers, regardless of programming experience and access to computing resources, which should greatly expand the diversity of projects using these datasets.

ForestPlotR could be used to replicate and strengthen tests of previously posed research questions and encourage the development of new research directions relating to stem-mapped forest stand data. The package allows a variety of spatially-explicit neighborhood metrics to be calculated with ease, and their variation across stands to be explored through visualizations. Consequently, it allows any researchers collecting data in stem-mapped forest stands to include neighborhood metrics as covariates in their models. This could lead to investigations of tree growth responses to climate change (Buechling et al., 2017) and competitive interactions (Fortunel et al., 2016) being replicated in a greater diversity of systems. Further, the inclusion of spatially-explicit neighborhood metrics could permit deeper investigation of spatial variation in soil microbial communities (Otsing et al., 2021), and improve the

accuracy of models predicting carbon storage dynamics (Ma et al., 2021; Martínez Cano et al., 2020). We also hope, that by encouraging a greater diversity of researchers to use stem-mapped forest stand data, ForestPlotR will pave the way for many novel and creative research questions to be asked with this data type.

This initial version of ForestPlotR is designed to facilitate what we expect to be the most common uses of stem-mapped forest stand data, but we plan to expand its functionality over time with input from the research community. Some potential expansions include: (1) a rapid-fitting tree survival model that uses regularized logistic regression, (2) the ability to extract average annual growth rates over a user-defined time period, (3) functions that connect to other packages or publicly available datasets to pull in stand-level climatic data (e.g. WorldClim data obtained through raster::getData) or tree-level topographical data (e.g. lidar). However, as the code underlying ForestPlotR is openly available on GitHub, we also hope that the growing community of researchers working with stem-mapped forest stand data will contribute improvements and additions that we have not considered.

The ForestPlotR package can be downloaded at https://github.com/sgraham9319/ForestPlotR and a website containing detailed vignettes for the common uses of the package is available at https://sgraham9319.github.io/ForestPlotR/index.html. Users of the package are encouraged to report difficulties and suggest improvements through pull requests and by posting issues on GitHub.

2.5. Acknowledgements

We thank Joe Ammirati, Michele Buonanduci, Tony Cannistra, Brian Harvey, Aji John, Rubén Delgado Manzanedo, Kavya Pradhan, Meera Lee Sethi, Hunter Stanke, Kristiina Visakorpi, the ETH Zürich Plant Ecology Group, and the UW eScience Incubator program for support and feedback throughout this project. The tree growth data included in the package were provided courtesy of the Pacific Northwest Permanent Sample Plot Program, in partnership with the HJ Andrews Experimental Forest and Long Term Ecological Research (LTER) program, which are administered cooperatively by the USDA Forest Service Pacific Northwest Research Station, Oregon State University, and the Willamette National Forest.

2.6. Data availability statement

The data used in the case study section are built-in to the ForestPlotR package, and the code underlying the case study is available at https://github.com/sgraham9319/ForestPlotR_case_study.

2.7. References

Buechling, A., Martin, P. H., & Canham, C. D. (2017). Climate and competition effects on tree growth in Rocky Mountain forests. *Journal of Ecology*, *105*(6), 1636–1647.

Buonanduci, M. S., Morris, J. E., Agne, M. C., & Harvey, B. J. (2020). Neighborhood context mediates probability of host tree mortality in a severe bark beetle outbreak. *Ecosphere*, *11*(8).

Condit, R. (1995). Research in large, long-term tropical forest plots. *Trends in Ecology & Evolution*, *10*(1), 18–22.

Contreras, M. A., Affleck, D., & Chung, W. (2011). Evaluating tree competition indices as predictors of basal area increment in western Montana forests. *Forest Ecology and Management*, *262*(11), 1939–1949.

Davies, S. J., Abiem, I., Abu Salim, K., Aguilar, S., Allen, D., Alonso, A., Anderson-Teixeira, K., Andrade, A., Arellano, G., Ashton, P. S., Baker, P. J., Baker, M. E., Baltzer, J. L., Basset, Y., Bissiengou, P., Bohlman, S., Bourg, N. A., Brockelman, W. Y., Bunyavejchewin, S., … Zuleta, D. (2021). ForestGEO: Understanding forest diversity and dynamics through a global observatory network. *Biological Conservation*, *253*, 108907.

Fortunel, C., Lasky, J. R., Uriarte, M., Valencia, R., Wright, S. J., Garwood, N. C., & Kraft, N. J. B. (2018). Topography and neighborhood crowding can interact to shape species growth and distribution in a diverse Amazonian forest. *Ecology*, *99*(10), 2272–2283.

Fortunel, C., Valencia, R., Wright, S. J., Garwood, N. C., & Kraft, N. J. B. (2016). Functional trait differences influence neighbourhood interactions in a hyperdiverse Amazonian forest. *Ecology Letters*, *19*(9), 1062–1070.

Franklin, J. F., Bell, D., Shaw, D. C. Long-Term Growth, Mortality and Regeneration of Trees in Permanent Vegetation Plots in the Pacific Northwest, 1910 to Present ver 18. Environmental Data Initiative. Available online: https://doi.org/10.6073/pasta/45a1f16b3d8ecd0585a2d2e115c07d41 (accessed on 14 September 2021).Graham, S. I., Rokem, A., Fortunel, C., Kraft, N. J. B., & Lambers, J. H. R. (2021). Regularized regression: A new tool for investigating and predicting tree growth. *Forests, 12*(9), 1283.

Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., Blomberg, S. P., & Webb, C. O. (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, *26*(11), 1463–1464.

Kunstler, G., Falster, D., Coomes, D. A., Hui, F., Kooyman, R. M., Laughlin, D. C., Poorter, L., Vanderwel, M., Vieilledent, G., Wright, S. J., Aiba, M., Baraloto, C., Caspersen, J., Cornelissen, J. H. C., Gourlet-Fleury, S., Hanewinkel, M., Herault, B., Kattge, J., Kurokawa, H., … Westoby, M. (2016). Plant functional traits have globally consistent effects on competition. *Nature*, *529*(7585), 204–207.

Laliberté, E., & Legendre, P. (2010). A distance-based framework for measuring functional diversity from multiple traits. *Ecology*, *91*(1), 299–305.

Laliberté, E., Legendre, P., & Shipley, B. (2014). FD: measuring functional diversity from multiple traits, and other tools for functional ecology. R package version 1.0-12.

Lepore, M., Arellano, G., Condit, R., Davies, S., Detto, M., Gonzalez-Akre, E., Hall, P., Harms, K., Herrmann, V., Kenfack, D., Lao, S., McMahon, S., Russo, S., Anderson-Teixeira, K., Zemunik, G., & Zuleta, D. (2019). fgeo: analyze forest diversity and dynamics. R package version 1.1.4.

Lutz, J. A. (2015). The evolution of long-term data for forestry: Large temperate research plots in an era of global change. *Northwest Science*, *89*(3), 255–269.

Ma, L., Hurtt, G., Tang, H., Lamb, R., Campbell, E., Dubayah, R., Guy, M., Huang, W., Lister, A., Lu, J., O'Neil-Dunne, J., Rudee, A., Shen, Q., & Silva, C. (2021). High-resolution forest carbon modelling for climate mitigation planning over the RGGI region, USA. *Environmental Research Letters*, *16*(4), 045014.

Martínez Cano, I., Shevliakova, E., Malyshev, S., Wright, S. J., Detto, M., Pacala, S. W., & Muller-Landau, H. C. (2020). Allometric constraints and competition enable the simulation of size structure and carbon fluxes in a dynamic vegetation model of tropical forests (LM3PPA-TV). *Global Change Biology*, *26*(8), 4478–4494.

Otsing, E., Anslan, S., Ambrosio, E., Koricheva, J., & Tedersoo, L. (2021). Tree species richness and neighborhood effects on ectomycorrhizal fungal richness and community structure in boreal forest. *Frontiers in Microbiology*, *12*, 567961.

Rouvinen, S., & Kuuluvainen, T. (1997). Structure and asymmetry of tree crowns in relation to local competition in a natural mature Scots pine forest. *Canadian Journal of Forest Research*, *27*, 890–902.

Stanke, H., Finley, A. O., Weed, A. S., Walters, B. F., & Domke, G. M. (2020). rFIA: An R package for estimation of forest attributes with the US Forest Inventory and Analysis database. *Environmental Modelling & Software*, *127*, 104664.

Tredennick, A. T., Hooker, G., Ellner, S. P., & Adler, P. B. (2021). A practical guide to selecting models for exploration, inference, and prediction in ecology. *Ecology*, *102*(6).

Uriarte, M., Canham, C. D., Thompson, J., & Zimmerman, J. K. (2004). A neighborhood analysis of tree growth and survival in a hurricane-driven tropical forest. *Ecological Monographs*, *74*(4), 591–614.

Wiegand, T., Uriarte, M., Kraft, N. J. B., Shen, G., Wang, X., & He, F. (2017). Spatially explicit metrics of species diversity, functional diversity, and phylogenetic diversity: Insights into plant community assembly processes. *Annual Review of Ecology, Evolution, and Systematics*, *48*(1), 329–351.

2.8. Figures

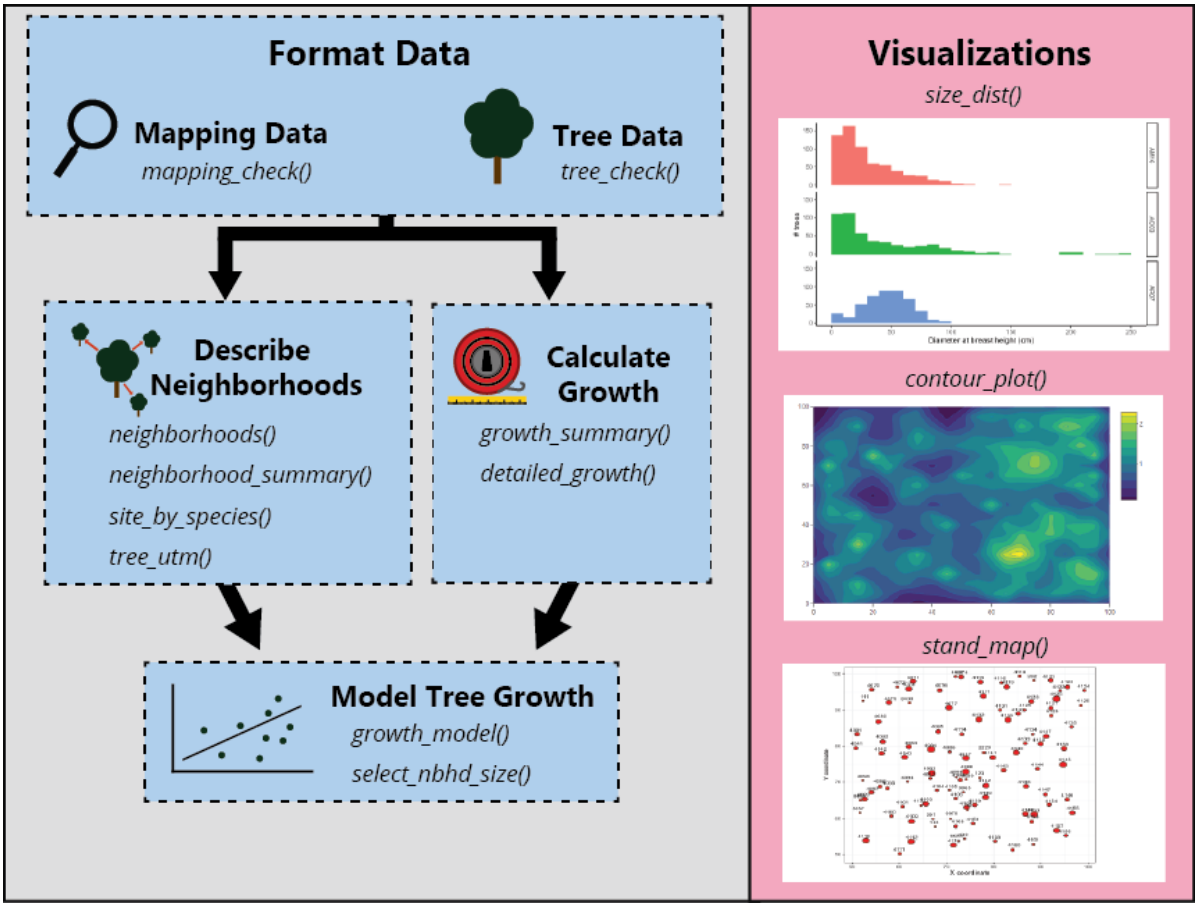

Figure 2.1. Schematic of ForestPlotR functions. Black arrows represent a potential analysis plan that combines multiple function categories; formatted mapping and tree measurement datasets can be used to obtain quantitative neighborhood descriptions and calculate tree growth rates, which can be combined to develop tree growth models. However, users who do not wish to develop such models may be more interested in using functions from the "Describe Neighborhoods" and "Visualizations" categories for exploratory analysis.

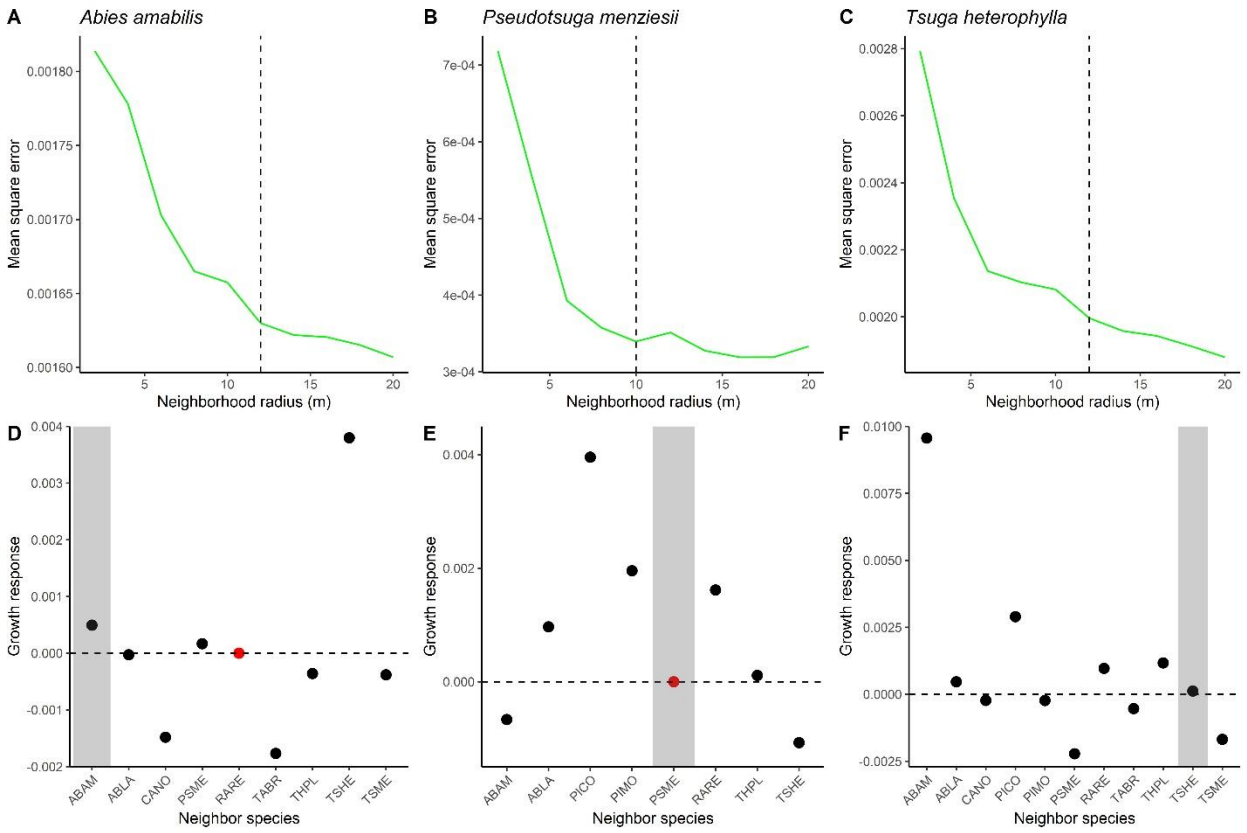

Figure 2.2. Using ForestPlotR to investigate how the growth rates of common tree species in Mount Rainier National Park are influenced by the species identity of neighboring trees. Panels A-C show the relationship between model fit (mean square error) and size of tree neighborhoods, with dashed vertical lines indicating the neighborhood size selected for further analysis. Plots D-F show the growth response of each focal species to neighbors of particular species, with conspecific neighbors shaded in gray and the points for neighbor species whose effect was dropped during model fitting colored in red. Plots are aligned according to the focal species they represent: A and D = *A. amabilis*, B and E = *P. menziesii*, C and F = *T. heterophylla*. Codes for competitor species are defined as follows: ABAM = A. amablis, ABLA = *Abies lasiocarpa*, CANO = *Callitropsis nootkatensis*, PICO = *Pinus contorta*, PIMO = *Pinus monitcola*, PSME = *P. menziesii*, RARE = all trees of species represented by < 100 interactions, TABR = *Taxus brevifolia*, THPL = *Thuja plicata*, TSHE = *T. heterophylla*, TSME = *Tsuga mertensiana*.

Chapter 3. Subalpine meadow encroachment by *Abies lasiocarpa* may be unimpeded by limited availability of ectomycorrhizal fungi

Stuart I. Graham, Lauren Dorsch, Joe Ammirati, Janneke Hille Ris Lambers

Abstract

Species interactions have been shown to influence the rate at which many species are shifting their ranges in response to climate change. The symbiotic interaction between ectomycorrhizal (EM) fungi and their plant hosts may be particularly important but has received little attention in the range shift literature thus far. Plants that depend on EM fungi may encounter few suitable fungal associates beyond their current range limits because of high host-fungus specificity and the surprisingly limited dispersal abilities of EM fungi, raising the question of whether the range shifts of these plants could be hindered by the availability of EM fungi. We investigated this question by studying the EM associations of *Abies lasiocarpa*, a treeline tree species predicted to encroach upon subalpine meadows devoid of EM host plants as climate change mediates an upslope range shift. We found that mature *A. lasiocarpa* that have recently colonized meadows had similar levels of EM colonization as conspecifics growing in the nearby forest, despite large differences in soil conditions known to influence EM fungal community composition. Moreover, a bioassay revealed that seedlings grown in soils sampled from meadows, with few or no mature conspecifics nearby, formed associations with a reduced diversity of EM fungi but achieved comparable levels of colonization and growth compared to seedlings grown in soil sampled from the nearby forest. Overall, our results suggest that dispersal limitation of EM fungi may reduce the diversity of suitable fungal symbionts available for EM host plants during range shifts, but that this may not impede the early establishment of EM host plant seedlings. It remains unclear whether the communities of EM fungal associates of range-shifting EM host plants diversify as the plants mature, and if so, how this fungal diversity arrives. To answer this question, future studies should characterize the EM fungal communities

associating with naturally-occurring EM plant hosts of all life stages growing beyond the traditional range limit.

3.1. Introduction

The redistribution of species to higher elevations and latitudes is one of the most widely documented impacts of climate change (Parmesan, 2006; Sunday et al., 2012). However, the rate of range shift differs dramatically among species (Chen et al., 2011), and understanding the drivers and ecological repercussions of this variation remain among the biggest challenges in contemporary ecology (Alexander et al., 2015; Paquette & Hargreaves, 2021; Pecl et al., 2017). Perhaps the most frequently implicated driver of complex range shift dynamics is species interactions (HilleRisLambers et al., 2013; Urban et al., 2013). Expansion at the leading range edge may be accelerated if a species encounters reduced negative interactions (e.g. release from natural enemies; Mackay et al., 2020) or increased positive interactions (e.g. more effective mutualists; Gundale et al., 2014) beyond its range limit. Alternatively, if a species experiences increased negative interactions (e.g. increased herbivory; Cairns & Moen, 2004) or reduced positive interactions (e.g. limited pollination services; Moeller et al., 2012), its rate of range expansion may be slowed.

One type of species interaction that has received relatively little attention in the range shift literature is the ectomycorrhizal (EM) symbiosis. Due to their high dependency on EM fungal associations, plant species that host EM fungi may find their range shifts moderated by the availability of suitable EM fungi beyond their current range limit. In the EM symbiosis, the fungus absorbs nutrients (mainly N and P) and moisture from the surrounding soil and transfers them to its plant host in exchange for photosynthate (Smith & Read, 2008). Ectomycorrhizal host plants are considered incapable of extracting enough nutrition from the poor soils in which they usually reside without the assistance of EM fungi (Smith & Read, 2008), and therefore may be unable to shift their range into areas that lack EM fungi. Moreover, because EM fungi are entirely dependent on their plant host for chemical energy (Lindahl & Tunlid, 2015), their spores will not germinate in the absence of a suitable host plant. Considering that most EM fungi also specialize on a single plant family or genus (Molina et al., 1992) and only 2% of plant species

form EM symbioses (Brundrett & Tedersoo, 2018), it is highly likely that an EM host plant will find few suitable EM fungi actively growing beyond its current range limit (Lankau & Keymer, 2016).

Of course, EM fungi may co-migrate with their host plants, but there are a number of potential barriers to this scenario. First, co-migration requires that EM fungal spores can successfully disperse beyond the current range boundary and survive in a dormant state until the seed of a suitable host plant arrives (Dickie et al., 2017). There is growing evidence that many EM fungi experience substantial dispersal limitation (Peay et al., 2010, 2012), with most spores being deposited centimeters to meters from the fruiting body (Peay et al., 2016). Additionally, the majority of EM fungal spores are unlikely to remain viable long enough to establish effective spore banks (Nara, 2009; Nguyen et al., 2012). Second, the presence of suitable host plants is not the only criterion for habitat suitability of EM fungi. Soil characteristics such as pH, moisture content, and nutrient content have been repeatedly shown to correlate with the composition of EM fungal communities (Arraiano-Castilho et al., 2021; Castaño et al., 2018; Truong et al., 2019), and it is therefore possible that different cues will trigger the range shifts of EM fungi compared to their host plants. In sum, there are many conceivable scenarios where EM host plants may encounter few suitable EM fungi beyond their current range limit.

Even if EM Fungi are available beyond the range limit of the host plant, EM community composition is likely to differ substantially from that within the host range, with potential implications for plant performance and range shift success. In nature, the roots of a single EM host plant are typically colonized by a diverse set of EM fungal taxa (Bahram et al., 2014; Jumpponen et al., 2010). Although there is a large amount of functional redundancy among EM fungi (Peay et al., 2016), taxa are known to differ in important functional traits such as the volume of soil that their hyphae can explore (Agerer, 2001) and their ability to extract mineral nutrients from organic matter (Lindahl et al., 2021). Moreover, the community of EM fungi actively associated with an EM host plant changes in a somewhat deterministic manner over the host plant's lifetime; seedlings are typically colonized by a small number of highly-infective 'pioneer' EM fungi that are either joined or replaced by other taxa that only occur on mature

plants (Boeraeve et al., 2021; Clemmensen et al., 2015). It remains an open question whether the reduced communities of EM fungi present beyond the range limits of EM host plants will contain sufficient functional diversity to meet the needs of invading host plants at both juvenile and mature life stages.

In this study, we investigate whether the climate change-mediated range shift of subalpine meadow encroachment by the EM host tree *Abies lasiocarpa* is likely to be impeded by limited availability of EM fungi. This is a particularly interesting system in which to explore this question because the subalpine meadows contain no likely EM host plants and are expected to differ greatly from adjacent forest in abiotic conditions that may influence EM fungal communities. Moreover, a small number of isolated mature *A. lasiocarpa* trees are already growing in the subalpine meadows, enabling investigation of the EM symbioses of these early invaders. Using this system, we test the following hypotheses: (1) subalpine meadows differ from nearby forest patches in edaphic variables that are likely to influence EM fungal community composition, (2) the extent and diversity of EM fungal colonization on both seedling and mature *A. lasiocarpa* is positively correlated with the abundance of nearby EM host plants, (3) growth rates of *A. lasiocarpa* seedlings will be positively correlated with the abundance of nearby EM host plants (as a result of differences in the level of EM fungal colonization).

3.2. Methods

3.2.1. *Study site*

This study was conducted in the subalpine zone of Mount Rainier National Park, WA, USA. The regional climate is temperate maritime, experiencing warm dry summers and cold wet winters. Most of the precipitation in this high elevation area arrives as snow, which typically covers the ground from late October to early July. The subalpine zone is a heterogeneous mosaic of forest patches and subalpine wildflower meadows. It begins at around 1550 m elevation, where meadows occupy the first substantial gaps in forest cover, and extends to around 2000 m elevation, where the final forest patches give way to the alpine zone.

All study sites were located within 100 m of hiking trails extending from the Paradise Visitor Center within the park. Subalpine tree communities in this area are dominated by *Abies lasiocarpa*, with *Tsuga mertensiana* and *Abies amabilis* being reasonably common but never constituting an entire forest patch in the absence of *A. lasiocarpa*. The plant communities of the subalpine meadows are highly diverse, but dominated by species that form associations with only arbuscular and ericoid mycorrhizal fungi. The only plants in these meadows known to form associations with ectomycorrhizal fungi are sedges, but these are easily identified in the field and were avoided during sampling.

Previous research in this system has suggested that climate warming may lead to the encroachment of subalpine meadows by forest trees (Ford & HilleRisLambers, 2020). In support of this, many meadows contain isolated mature *A. lasiocarpa* individuals, sometimes with small *A. lasiocarpa* seedlings growing nearby. We consider these trees to represent early invaders in a climate warming-mediated range shift of *A. lasiocarpa*.

3.2.2. *Field sampling*

We identified 25 individuals or small clusters (max. three individuals) of mature *A. lasiocarpa* growing in 25 distinct subalpine meadows, at least 10 m from any other mature trees (hereafter, "meadow trees"). We conservatively defined trees as mature if they were at least 1 m in height because no trees less than 1.5 m in height were observed to produce cones. We paired each meadow tree with an *A. lasiocarpa* tree in the nearest forest patch that was of similar size and growing at a similar elevation, slope and aspect (hereafter, "forest trees"). This resulted in a total of 25 tree pairs, each containing a meadow tree and a forest tree, ranging in elevation from 1586 – 1810 m.

Soil samples were taken on four separate occasions for four distinct purposes (Table 3.1). At the first three soil sampling events, a trowel was used to collect an approximately 10 x 10 x 10 cm block of surface soil at the focal tree's dripline in each of the four cardinal directions. The four soil blocks for each meadow and forest tree were immediately pooled. This resulted in 50 pooled soil samples per sampling

event (i.e. one for each meadow and forest tree). At the fourth soil sampling event, where soil was collected for the bioassay, only nine meadow/forest tree pairs were sampled to keep sample size manageable. In addition, a soil sample was taken in each of these nine meadows at a location at least 10 m from any conifer trees (including the meadow tree and seedlings) to represent a location with no ectomycorrhizal host plants (hereafter, "meadow-no-tree soil"). To avoid creating soil samples with artificially high ectomycorrhizal diversity, only a single soil block was taken for each bioassay soil sample rather than pooling four separate samples. This resulted in a total of 3 sampling locations x 9 forest/meadow pairs = 27 soil samples for the bioassay. During all soil sampling events, soil sampling tools were sterilized with 10% bleach between samples to prevent cross-contamination.

All collected soil samples were transported to the University of Washington on ice. The soil samples used for measuring soil chemistry variables (i.e. the first and third sampling events; Table 3.1) were dried for 48 hours before being put through a 2 mm sieve and stored at room temperature. A 2 mm sieve was used to extract *A. lasiocarpa* roots from the soil samples intended for measuring the extent of mature tree ectomycorrhizal colonization. The roots were then washed with de-ionized water and stored in 60 % ethanol at 4 °C. The soil samples used for the bioassay were put through a 2 mm sieve and stored at 4 °C. During the processing of all soil samples, tools that came into contact with soil were sterilized with 60 % ethanol between samples to prevent cross-contamination.

3.2.3. *Soil chemistry*

To test our first hypothesis, that meadow and forest soil conditions differ in a way that may influence EM fungal communities, we compared six aspects of soil chemistry between forest tree and meadows tree soils. Tests for measuring plant-available phosphorus (PO_4), plant-available nitrogen (NH_4 and NO_3), and C:N ratio of soil samples were conducted by the Soil Analytics Lab of the University of Washington. Water holding capacity was measured as the mass of the water contained in saturated soil as a percentage of the mass of the saturated soil. To measure pH, 5 g of soil were combined with 10 ml de-ionized water to create a solution before using a pH meter. Soil samples from only 10 of the meadow/forest tree pairs

were used for plant-available nitrogen tests due to funding constraints, but measurements of other soil chemistry variables were obtained for all 25 meadow/forest tree pairs.

3.2.4. *Bioassay*

We conducted a bioassay experiment to test our second and third hypotheses that the extent and diversity of EM fungal associations and growth of *A. lasiocarpa* seedlings would be positively correlated with the abundance of nearby EM host plants. Seedlings were grown in three soil types: forest tree soils (many nearby EM host plants), meadow tree soils (few nearby EM host plants), and meadow-no-tree soils (no nearby EM host plants; see *Field sampling* for details). We expected that the extent and diversity of EM fungal colonization, and seedling growth, would be highest for seedlings grown in forest tree soil and lowest for seedlings grown in meadow-no-tree soil.

We obtained *A. lasiocarpa* seed from Silvaseed Company (Roy, WA, USA). Seeds were soaked in de-ionized water for 24 hours, surface-sterilized with 10% bleach, and then stratified at 4 °C for 50 days to break seed dormancy. We planted 2000 *A. lasiocarpa* seeds in germination flats containing potting soil that had been sterilized through autoclaving (2 cycles of 1 hour at 121 °C). Seedling flats were placed inside a growth chamber that maintained a 16h: 8h light (24 °C): dark (21 °C) cycle and watered with de-ionized water every 1-2 days. After 2 months, approximately 100 of the 2000 seeds had germinated.

The three soil types used in the experiment (forest tree, meadow tree, and meadow-no-tree) were each sampled from nine meadow/forest tree pairs, and each combination was replicated three times, resulting in a total of 81 experimental seedlings. Of the three replicates of each soil type x meadow/forest tree pair location, one was selected as a control and given only sterilized soil to evaluate whether we were successful in preventing ectomycorrhizal contamination of our seedlings from sources other than the provided soil (e.g. airborne spores). Each of the 81 seedlings was transplanted from the germination flats into a conetainer (1.5 inches diameter x 8.25 inches deep; Stuewe & Sons, Tangent, OR, USA) that contained a small wad of pillow stuffing for drainage and a 2:2:1 mixture of the relevant experimental

field soil, sphagnum peat moss, and sand, respectively. The sphagnum peat moss and sand (and also the experimental field soil for the control seedlings) were sterilized through autoclaving (2 cycles of 1 hour at 121 °C). Conetainers were kept in the same growth chamber conditions as the germination flats and watered with de-ionized water every 2-4 days.

Seedlings were harvested 115 days after the transplant. Aboveground tissue of each seedling was dried for 48 hours and then weighed to measure aboveground biomass for the evaluation of seedling growth. The root system of each seedling was inspected visually under a dissecting microscope to look for ectomycorrhizal colonization. The colonized root-tips of each seedling were categorized according to their morphology and 1-2 samples of each morphological type per seedling were placed in de-ionized water and frozen at -80 °C for later DNA extraction (to evaluate the diversity of EM fungal colonization). The remaining root system (or a random portion of the root system if it contained > 300 root-tips) was put in 60% ethanol and stored at 4 °C for later calculations of the extent of mycorrhizal colonization.

3.2.5. *Percent colonization*

To test part of our second hypothesis, that the extent of EM fungal colonization would be positively correlated with the abundance of nearby EM host plants, we estimated the percentage of root-tips colonized for mature *A. lasiocarpa* in the field and our bioassay seedlings. We expected the mature forest trees would have higher colonization than mature meadow trees, and that seedlings grown in forest tree soil would have the highest colonization levels among bioassay seedlings. For both mature tree and seedling samples, root-tips were inspected under a dissecting microscope and classified as colonized or uncolonized based on the presence of root hairs, shape, texture, and color. We then calculated the percentage of root-tips that were colonized. To validate this method, we evaluated the consistency of findings by having two independent observers assess a subset of the samples (seven bioassay seedlings) and comparing their results.

3.2.6. *Molecular analysis*

To test another part of our second hypothesis, that the diversity of EM fungal associations would be positively correlated with the abundance of nearby EM host plants, we identified the EM fungal taxa observed on bioassay seedling roots through a molecular barcoding approach. We extracted and amplified fungal DNA from frozen seedling root-tips using the REDExtract-N-Amp Plant PCR Kit (Sigma-Aldrich, St. Louis, MO, USA). Specifically, a 1-2 mm section of each colonized root-tip was placed in 10 µl of the extraction solution from the PCR kit. The sample was then run through the thermocycler (25 °C for 10 minutes, 85 °C for 10 minutes) before adding 30 µl of the neutralization solution to stop the reaction. We then amplified the internal transcribed spacer (ITS) region using the primers ITS1F (Gardes & Bruns, 1993) and ITS4 (White et al., 1990), each at a concentration of 10 µM. Each PCR reaction contained 10 µl REDExtract-N-Amp PCR Reaction Mix, 1 µl ITS1F primer, 1 µl ITS4 primer, 2 µl DNA template, and 6 µl sterile water, for a total volume of 20 µl. The PCR thermocycler protocol was: (1) 95 °C for 1 minute, (2) 95 °C for 30 seconds, (3) 55 °C for 20 seconds, (4) 72 °C for 50 seconds, (5) return to step (2) 34 times, (6) 72 °C for 5 minutes. Amplified DNA was cleaned using the ExoSAP enzymatic clean-up kit (Applied Biosystems, Waltham, MA, USA), where 5 µl of diluted ExoSAP (1 part ExoSAP to 3 parts sterile water) was combined with 8.75 µl PCR product and run in the thermocycler (37 °C for 45 minutes, 85 °C for 15 minutes). Cleaned PCR products were sent to Genewiz, Inc. (Seattle, WA, USA) for Sanger sequencing. The obtained ITS sequences were trimmed and matched to sequences using the NCBI Nucleotide BLAST tool with default conditions. The top named BLAST hit for each sequence is reported if the percent identity was ≥ 97%. The communities of EM fungi observed on seedlings grown in each of the three soil types, which differed in the abundance of nearby EM host plants, were then compared.

3.3. Results

3.3.1. *Soil Chemistry*

Soil sampled from the rhizospheres of meadow trees had a higher concentration of plant-available phosphorus, higher pH, and a lower C:N ratio than soil sampled from the rhizospheres of paired forest trees (Figure 3.1). However, there was no difference in the concentration of plant-available nitrogen in the

form of either ammonium or nitrate. Although meadow soils tended to have a higher water holding capacity than their paired forest soils, this difference was not statistically significant after a Bonferroni correction for multiple testing (critical threshold for p-value = 0.05 / 6 soil tests = 0.008).

3.3.2. *Bioassay*

Some of the control seedlings (i.e. grown in sterilized soil) had low levels of mycorrhizal colonization (4.3 ± 4.5% (mean ± se) root-tips colonized), but this was far lower than the levels observed for experimental seedlings (i.e. grown in live soil; 30.9 ± 5.6%; t-test, $p < 0.001$). Experimental seedlings also tended to have higher biomass (0.10 ± 0.01g) than control seedlings (0.08 ± 0.01g), but this difference was not statistically significant (t-test; $p = 0.069$). The two independent observers obtained similar estimates for the extent of mycorrhizal colonization on the seven seedlings they both evaluated (Pearson's correlation coefficient = 0.91; $p = 0.005$), thereby validating this methodology. Among the experimental seedlings, there was no effect of soil source on the extent of mycorrhizal colonization or biomass (Figure 3.2).

We found a greater number of distinct fungal species on the roots of seedlings grown in forest tree soil, than those grown in both meadow tree and meadow-no-tree soils (Table 3.2). The Ascomycete ectomycorrhizal fungus, *Wilcoxina mikolae*, was the dominant taxon found in seedlings grown in both meadow tree and meadow-no-tree soil. The Ascomycete ectomycorrhizal fungus, *Cenococcum geophilum*, was found on seedlings grown in soils sampled from the rhizospheres of mature conspecifics (forest tree and meadow tree soils) but not on seedlings grown in soils with no nearby conspecifics (meadow-no-tree soil). The dark-septate fungus, *Phialocephala fortinii*, was found only on a single seedling grown in meadow-no-tree soil.

3.3.3. *Mature trees*

Mature *A. lasiocarpa* trees growing in the forest and in subalpine meadows did not differ in their extent of mycorrhizal colonization (Figure 3.3). In both environments, almost all root-tips appeared to be colonized by ectomycorrhizal fungi.

3.4. Discussion

Overall, we did not find any evidence to suggest that the encroachment of subalpine meadows by *A. lasiocarpa* will be impeded by limited availability of suitable EM fungi beyond the current range limit of *A. lasiocarpa* (i.e. treeline). In support of our first hypothesis, forest patch and subalpine meadow soils do differ along axes frequently reported to influence EM fungal community composition (Figure 3.1). However, mature *A. lasiocarpa* trees growing in the two environments had equal levels of EM fungal colonization (Figure 3.3). Moreover, seedlings grown in soil sampled from forest patches, meadows with nearby mature conspecifics, and meadows without nearby mature conspecifics, all achieved similar levels of EM fungal colonization and growth (Figure 3.2). There was some suggestion that a lower diversity of EM fungal taxa were available to colonize seedlings in the meadows (regardless of the presence of nearby mature conspecifics; Table 3.2), giving weak and only partial support for our second hypothesis, but because this did not translate into seedling performance differences we document no support for our third hypothesis. Together, these results are consistent with the idea that *A. lasiocarpa* encounters a reduced diversity of EM fungi in meadows relative to inside its current range, but this may not impede the initial establishment of *A. lasiocarpa* seedlings in these meadows. We next discuss the caveats of these conclusions and their implications for how the EM symbiosis may influence the range shifts of EM host plants.

Our results are consistent with the idea that EM fungi experience substantial dispersal limitation (Peay et al., 2010, 2012). A total of six distinct EM fungal taxa colonized seedlings grown in forest soil, yet only a subset of these were found on seedlings grown in meadow and meadow-no-tree soil (except for *Phialocephala fortinii*, which was unique to seedlings grown in meadow-no-tree soil; Table 3.2). This could indicate that the taxa found only on seedlings grown in forest soil (*Amanita muscaria*, *Hyaloscypha*

variabilis, Helotiaceae sp.) were unable to disperse their spores in meaningful quantities across the mere tens of meters separating forest and meadow sampling locations. Moreover, the EM fungus *Wilcoxina mikolae*, that colonized the majority of seedlings grown in meadow tree and meadow-no-tree soil, is one of the few EM fungi known to have sufficient spore longevity to develop a viable soil spore bank (Nguyen et al., 2012; Policelli et al., 2019). *Cenococcum geophilum* was observed only on the roots of seedlings grown in soil sampled close to a mature conspecific (i.e. forest tree and meadow tree soils), which is consistent with a previous study that found field transplanted *Pseudotsuga menziesii* saplings only formed associations with *C. geophilum* when planted close to the forest edge (Grove et al., 2019). Although we cannot discount the role of stochasticity in our small experiment, these intriguing patterns suggest that dispersal limitation of EM fungi could play a large role in determining the EM symbioses that EM host plants will form during range shifts.

The role of EM symbioses in constraining EM host plant range shifts is likely to be small if differences in the identity of EM fungal associates do not impact performance of the earliest colonizers. Our results suggest that, at least at the seedling stage, soil type did not influence overall EM colonization (even as it influenced diversity) and that seedlings achieved similar biomass (Figure 3.2) despite often being colonized by different EM fungal taxa. Many studies have demonstrated the benefits of EM fungal colonization for EM host plant seedling performance (Hawkins et al., 2015; Jenkins et al., 2018) and our finding of low colonization and somewhat reduced biomass in the control seedlings supports this. However, very little is known about how variation in the identity of EM fungal associates influences host plant performance. It is generally assumed that EM host plants benefit from a diversity of EM fungal associates due to both the sampling effect and some degree of functional complementarity (Diagne et al., 2013; Kipfer et al., 2012) but there also appears to be a high degree of functional redundancy among EM fungi (Peay et al., 2016). It therefore seems plausible that in the earliest life stages of an EM host plant, many EM fungi are functionally equivalent. However, it is also likely that any impacts of functional differences in EM fungal associates would be more apparent under harsher field conditions, where the

added benefits of EM fungal colonization, such as pathogen protection (Marx, 1972) and drought resilience (Sebastiana et al., 2018), may be more important. If the identity of EM fungal associates has little impact on seedling performance, range shifts may be initially unaffected by limited EM fungal diversity, but this may represent a barrier as the seedlings mature.

The establishment of EM host plant seedlings beyond their range limit may be influenced by additional benefits of nearby mature conspecifics that were not accounted for in the bioassay experiment. In nature, an EM host plant can become colonized by a new EM fungus either by triggering the germination of a fungal spore or by joining an established mycelial network (Ishida et al., 2008). In the bioassay experiment, however, only the spore germination method is possible because any mycelial tissue would have died when it was severed from its host plant(s) during soil sampling. Importantly, a plant that can form an association with an established network of a particular EM fungus cannot necessarily trigger germination of the spores of that same EM fungus (Lofgren et al., 2018). Moreover, mycelial colonization can result in greater benefits for the host plant, including enhanced survival and access to nutrition obtained from other plants connected to the mycelial network (Simard et al., 1997; Teste et al., 2009). Mature trees may also interact positively with seedlings by buffering them from stressful abiotic conditions (Bertness & Callaway, 1994), such as the extreme snowfall in our study system. In fact, these positive feedbacks, where plants alter the environment in a manner that benefits conspecifics, are widely reported among EM host plants (Bennett et al., 2017). In summary, the forest tree and meadow tree soil treatments in our bioassay experiment are likely missing many possible seedling performance gains that could result from the presence of mature adult *A. lasiocarpa* trees in meadows. By contrast, the seedlings grown in meadow-no-tree soil accurately represent the initial stage of the range shift, where EM fungal colonization depends on spores, which we know is possible due to the mature EM colonized trees already present in the meadows (Figure 3.3). However, it seems likely that the established meadow trees will act as nuclei of further meadow encroachment by creating areas of more suitable habitat for conspecific seedlings.

Our finding of reduced diversity of EM fungal associations on seedlings grown in meadow tree soils raises interesting questions regarding the EM fungal associations of the mature *A. lasiocarpa* already growing in the subalpine meadows. In the similar high elevation treeline advancement of *Abies pinsapo* in Europe, the distributions of many EM fungi were observed to lag behind that of their host plant (Álvarez-Garrido et al., 2019). It is therefore possible that the mature *A. lasiocarpa* growing in meadows have a much reduced diversity of EM associates relative to their conspecifics in forest patches but are still performing well, as has been seen for *Abies alba* (Rudawska et al., 2016). However, the morphological variation we observed among colonized root-tips of mature *A. lasiocarpa* indicated a far greater diversity of EM fungi than were seen on the roots of bioassay seedlings. This added fungal diversity could suggest that far more EM fungal taxa are capable of forming persistent soil spore banks than is expected (Policelli et al., 2019), and these taxa were not observed in the bioassay experiment because their germination can only be triggered by adult *A. lasiocarpa* and not seedlings. Alternatively, these EM fungi could be growing on unexpected host plants, enabling them to colonize *A. lasiocarpa* from their established mycelium. Indeed, compatible EM fungi and pathogens of *Pinus cembra* have been observed above its high elevation range limit (Merges et al., 2018), and may be growing in association with Ericaceous shrubs such as *Arctostaphylos uva-ursi*, that are known to form arbutoid mycorrhizas with EM fungi (Krpata et al., 2007; Molina & Trappe, 1982; Zak, 1976). A third alternative we propose is that *A. lasiocarpa* seedlings are relatively indifferent to the identity of their first EM fungal associates, but that the formation of an EM symbiosis affords the seedlings time to wait for the rare dispersal of other EM fungal taxa with limited spore longevity from nearby forest patches. Widespread field sampling of the EM fungal communities of bulk soils and *A. lasiocarpa* of all life stages (using high-throughput sequencing approaches) is needed to investigate these hypotheses further.

In summary, this study suggests that the climate change-mediated range shift of a plant that depends on EM fungal symbionts may not be hindered by limited availability of suitable fungal symbionts beyond the current range boundary. It appears that at least some EM fungi are able to co-migrate with their plant host,

either by maintaining a persistent soil spore bank or by dispersing once host plant seedlings are established (Dickie et al., 2017). The ability of EM host plants to establish in areas of reduced EM fungal diversity (Lankau & Keymer, 2016) speaks to the widespread functional redundancy among EM fungi (Peay et al., 2016). Although it seems logical that a strong dependency on other species would only hinder a species' range shift (Moeller et al., 2012), empirical evidence for successful co-migration of EM fungi and their plant hosts is accumulating (Dickie et al., 2017; Moyano et al., 2020). However, the relative importance of persistent spore banks versus serendipitous dispersal in these co-migrations remains unclear. Future research on the influence of EM fungal availability on EM host plant range shifts should focus on comparing the communities of EM fungi associated with naturally-occurring seedlings and mature trees beyond the range limit using high-throughput sequencing approaches.

3.5. Acknowledgements

We would like to thank Aji John, Rubén Manzanedo, Kavya Pradhan, and Meera Lee Sethi for feedback on project development, Cole Lysgaard and Ben Simpson for assistance with field sampling, Aaron Hernández and Keith Possee for assistance with the bioassay experiment, and Itzue Caviedes Solis, Peter Kennedy, Ben Kerr, and Olivia Kosterlitz for assistance with molecular methods. Funding for this project came from Oregon Mycological Society, Mazamas, Mycological Society of America, Stuntz Mycology Fund, and the University of Washington Biology Department.

3.6. References

Agerer, R. (2001). Exploration types of ectomycorrhizae. *Mycorrhiza*, *11*(2), 107–114.

Alexander, J. M., Diez, J. M., & Levine, J. M. (2015). Novel competitors shape species' responses to climate change. *Nature*, *525*(7570), 515–518.

Álvarez-Garrido, L., Viñegla, B., Hortal, S., Powell, J. R., & Carreira, J. A. (2019). Distributional shifts in ectomycorrizhal fungal communities lag behind climate-driven tree upward migration in a conifer forest-high elevation shrubland ecotone. *Soil Biology and Biochemistry*, *137*, 107545.

Arraiano-Castilho, R., Bidartondo, M. I., Niskanen, T., Clarkson, J. J., Brunner, I., Zimmermann, S., Senn-Irlet, B., Frey, B., Peintner, U., Mrak, T., & Suz, L. M. (2021). Habitat specialisation controls ectomycorrhizal fungi above the treeline in the European Alps. *New Phytologist*, *229*(5), 2901–2916.

Bahram, M., Harend, H., & Tedersoo, L. (2014). Network perspectives of ectomycorrhizal associations. *Fungal Ecology*, *7*, 70–77.

Bennett, J. A., Maherali, H., Reinhart, K. O., Lekberg, Y., Hart, M. M., & Klironomos, J. (2017). Plant-soil feedbacks and mycorrhizal type influence temperate forest population dynamics. *Science*, *355*(6321), 181–184.

Bertness, M. D., & Callaway, R. (1994). Positive interactions in communities. *Trends in Ecology & Evolution*, *9*(5), 191–193.

Boeraeve, M., Everts, T., Vandekerkhove, K., De Keersmaeker, L., Van de Kerckhove, P., & Jacquemyn, H. (2021). Partner turnover and changes in ectomycorrhizal fungal communities during the early life stages of European beech (Fagus sylvatica L.). *Mycorrhiza*, *31*(1), 43–53.

Brundrett, M. C., & Tedersoo, L. (2018). Evolutionary history of mycorrhizal symbioses and global host plant diversity. *New Phytologist*, *220*(4), 1108–1115.

Cairns, D. M., & Moen, J. (2004). Herbivory influences tree lines. *Journal of Ecology*, *92*(6), 1019–1024.

Castaño, C., Lindahl, B. D., Alday, J. G., Hagenbo, A., Martínez de Aragón, J., Parladé, J., Pera, J., & Bonet, J. A. (2018). Soil microclimate changes affect soil fungal communities in a Mediterranean pine forest. *New Phytologist*, *220*(4), 1211–1221.

Chen, I.-C., Hill, J. K., Ohlemüller, R., Roy, D. B., & Thomas, C. D. (2011). Rapid Range Shifts of Species Associated with High Levels of Climate Warming. *Science*, *333*(6045), 1024–1026.

Clemmensen, K. E., Finlay, R. D., Dahlberg, A., Stenlid, J., Wardle, D. A., & Lindahl, B. D. (2015). Carbon sequestration is related to mycorrhizal fungal community shifts during long-term succession in boreal forests. *New Phytologist*, *205*(4), 1525–1536.

Diagne, N., Thioulouse, J., Sanguin, H., Prin, Y., Krasova-Wade, T., Sylla, S., Galiana, A., Baudoin, E., Neyra, M., Svistoonoff, S., Lebrun, M., & Duponnois, R. (2013). Ectomycorrhizal diversity enhances growth and nitrogen fixation of Acacia mangium seedlings. *Soil Biology and Biochemistry*, *57*, 468–476.

Dickie, I. A., Bufford, J. L., Cobb, R. C., Desprez-Loustau, M., Grelet, G., Hulme, P. E., Klironomos, J., Makiola, A., Nuñez, M. A., Pringle, A., Thrall, P. H., Tourtellot, S. G., Waller, L., & Williams, N. M. (2017). The emerging science of linked plant–fungal invasions. *New Phytologist*, *215*(4), 1314–1332.

Ford, K. R., & HilleRisLambers, J. (2020). Soil alters seedling establishment responses to climate. *Ecology Letters*, *23*(1), 140–148.

Gardes, M., & Bruns, T. D. (1993). ITS primers with enhanced specificity for basidiomycetes—Application to the identification of mycorrhizae and rusts. *Molecular Ecology*, *2*(2), 113–118.

Grove, S., Saarman, N. P., Gilbert, G. S., Faircloth, B., Haubensak, K. A., & Parker, I. M. (2019). Ectomycorrhizas and tree seedling establishment are strongly influenced by forest edge proximity but not soil inoculum. *Ecological Applications*, *29*(3), e01867.

Gundale, M. J., Kardol, P., Nilsson, M., Nilsson, U., Lucas, R. W., & Wardle, D. A. (2014). Interactions with soil biota shift from negative to positive when a tree species is moved outside its native range. *New Phytologist*, *202*(2), 415–421.

Hawkins, B. J., Jones, M. D., & Kranabetter, J. M. (2015). Ectomycorrhizae and tree seedling nitrogen nutrition in forest restoration. *New Forests*, *46*(5–6), 747–771.

HilleRisLambers, J., Harsch, M. A., Ettinger, A. K., Ford, K. R., & Theobald, E. J. (2013). How will biotic interactions influence climate change-induced range shifts? *Annals of the New York Academy of Sciences*, *1297*, 112–125.

Ishida, T. A., Nara, K., Tanaka, M., Kinoshita, A., & Hogetsu, T. (2008). Germination and infectivity of ectomycorrhizal fungal spores in relation to their ecological traits during primary succession. *New Phytologist*, *180*(2), 491–500.

Jenkins, M. L., Cripps, C. L., & Gains-Germain, L. (2018). Scorched Earth: Suillus colonization of Pinus albicaulis seedlings planted in wildfire-impacted soil affects seedling biomass, foliar nutrient content, and isotope signatures. *Plant and Soil*, *425*(1–2), 113–131.

Jumpponen, A., Jones, K. L., David Mattox, J., & Yaege, C. (2010). Massively parallel 454-sequencing of fungal communities in *Quercus* spp. Ectomycorrhizas indicates seasonal dynamics in urban and rural sites. *Molecular Ecology*, *19*, 41–53.

Kipfer, T., Wohlgemuth, T., van der Heijden, M. G. A., Ghazoul, J., & Egli, S. (2012). Growth Response of Drought-Stressed Pinus sylvestris Seedlings to Single- and Multi-Species Inoculation with Ectomycorrhizal Fungi. *PLoS ONE*, *7*(4), e35275.

Krpata, D., Mühlmann, O., Kuhnert, R., Ladurner, H., Göbl, F., & Peintner, U. (2007). High diversity of ectomycorrhizal fungi associated with Arctostaphylos uva-ursi in subalpine and alpine zones: Potential inoculum for afforestation. *Forest Ecology and Management*, *250*(3), 167–175.

Lankau, R. A., & Keymer, D. P. (2016). Ectomycorrhizal fungal richness declines towards the host species' range edge. *Molecular Ecology*, *25*(13), 3224–3241.

Lindahl, B. D., Kyaschenko, J., Varenius, K., Clemmensen, K. E., Dahlberg, A., Karltun, E., & Stendahl, J. (2021). A group of ectomycorrhizal fungi restricts organic matter accumulation in boreal forest. *Ecology Letters*, *24*(7), 1341–1351.

Lindahl, B. D., & Tunlid, A. (2015). Ectomycorrhizal fungi – potential organic matter decomposers, yet not saprotrophs. *New Phytologist*, *205*(4), 1443–1447.

Lofgren, L., Nguyen, N. H., & Kennedy, P. G. (2018). Ectomycorrhizal host specificity in a changing world: Can legacy effects explain anomalous current associations? *New Phytologist*, *220*(4), 1273–1284.

Mackay, K. D., Gross, C. L., & Ryder, D. S. (2020). Increased reproductive success through parasitoid release at a range margin: Implications for range shifts induced by climate change. *Journal of Biogeography*, *47*(5), 1041–1055.

Marx, D. H. (1972). Ectomycorrhizae as Biological Deterrents to Pathogenic Root Infections. *Annual Review of Phytopathology*, *10*(1), 429–454.

Merges, D., Bálint, M., Schmitt, I., Böhning-Gaese, K., & Neuschulz, E. L. (2018). Spatial patterns of pathogenic and mutualistic fungi across the elevational range of a host plant. *Journal of Ecology*, *106*(4), 1545–1557.

Moeller, D. A., Geber, M. A., Eckhart, V. M., & Tiffin, P. (2012). Reduced pollinator service and elevated pollen limitation at the geographic range limit of an annual plant. *Ecology*, *93*(5), 1036–1048.

Molina, R., Massicotte, H., & Trappe, J. (1992). *Specificity phenomena in mycorrhizal symbioses: Community-ecological consequences and practical implications*. Chapman and Hall.

Molina, R., & Trappe, J. M. (1982). Lack of mycorrhizal specificity by the ericaceous hosts Arbutus menziesii and Arctostaphylos uva-ursi. *New Phytologist*, *90*, 495–509.

Moyano, J., Rodriguez-Cabal, M. A., & Nuñez, M. A. (2020). Highly invasive tree species are more dependent on mutualisms. *Ecology*, *101*(5), e02997.

Nara, K. (2009). Spores of ectomycorrhizal fungi: Ecological strategies for germination and dormancy. *New Phytologist*, *181*(2), 245–248.

Nguyen, N. H., Hynson, N. A., & Bruns, T. D. (2012). Stayin' alive: Survival of mycorrhizal fungal propagules from 6-yr-old forest soil. *Fungal Ecology*, *5*(6), 741–746.

Paquette, A., & Hargreaves, A. L. (2021). Biotic interactions are more often important at species' warm versus cool range edges. *Ecology Letters*, *24*(11), 2427–2438.

Parmesan, C. (2006). Ecological and Evolutionary Responses to Recent Climate Change. *Annual Review of Ecology, Evolution, and Systematics*, *37*(1), 637–669.

Peay, K. G., Garbelotto, M., & Bruns, T. D. (2010). Evidence of dispersal limitation in soil microorganisms: Isolation reduces species richness on mycorrhizal tree islands. *Ecology*, *91*(12), 3631–3640.

Peay, K. G., Kennedy, P. G., & Talbot, J. M. (2016). Dimensions of biodiversity in the Earth mycobiome. *Nature Reviews Microbiology*, *14*(7), 434–447.

Peay, K. G., Schubert, M. G., Nguyen, N. H., & Bruns, T. D. (2012). Measuring ectomycorrhizal fungal dispersal: Macroecological patterns driven by microscopic propagules. *Molecular Ecology*, *21*(16), 4122–4136.

Pecl, G. T., Araújo, M. B., Bell, J. D., Blanchard, J., Bonebrake, T. C., Chen, I.-C., Clark, T. D., Colwell, R. K., Danielsen, F., Evengård, B., Falconi, L., Ferrier, S., Frusher, S., Garcia, R. A., Griffis, R. B., Hobday, A. J., Janion-Scheepers, C., Jarzyna, M. A., Jennings, S., … Williams, S. E. (2017).

Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science*, *355*(6332), eaai9214.

Policelli, N., Bruns, T. D., Vilgalys, R., & Nuñez, M. A. (2019). Suilloid fungi as global drivers of pine invasions. *New Phytologist*, *222*(2), 714–725.

Rudawska, M., Pietras, M., Smutek, I., Strzeliński, P., & Leski, T. (2016). Ectomycorrhizal fungal assemblages of Abies alba Mill. Outside its native range in Poland. *Mycorrhiza*, *26*(1), 57–65.

Sebastiana, M., da Silva, A. B., Matos, A. R., Alcântara, A., Silvestre, S., & Malhó, R. (2018). Ectomycorrhizal inoculation with Pisolithus tinctorius reduces stress induced by drought in cork oak. *Mycorrhiza*, *28*(3), 247–258.

Simard, S. W., Perry, D. A., Jones, M. D., Myrold, D. D., Durall, D. M., & Molina, R. (1997). Net transfer of carbon between ectomycorrhizal tree species in the field. *Nature*, *388*(6642), 579–582.

Smith, S., & Read, D. (2008). *Mycorrhizal symbiosis* (3rd ed.). Academic Press.

Sunday, J. M., Bates, A. E., & Dulvy, N. K. (2012). Thermal tolerance and the global redistribution of animals. *Nature Climate Change*, *2*(9), 686–690.

Teste, F. P., Simard, S. W., Durall, D. M., Guy, R. D., Jones, M. D., & Schoonmaker, A. L. (2009). Access to mycorrhizal networks and roots of trees: Importance for seedling survival and resource transfer. *Ecology*, *90*(10), 2808–2822.

Truong, C., Gabbarini, L. A., Corrales, A., Mujic, A. B., Escobar, J. M., Moretto, A., & Smith, M. E. (2019). Ectomycorrhizal fungi and soil enzymes exhibit contrasting patterns along elevation gradients in southern Patagonia. *New Phytologist*, *222*(4), 1936–1950.

Urban, M. C., Zarnetske, P. L., & Skelly, D. K. (2013). Moving forward: Dispersal and species interactions determine biotic responses to climate change: Dispersal and species interactions. *Annals of the New York Academy of Sciences*, *1297*, 44–60.

White, T. J., Bruns, T. D., Lee, S. B., & Taylor, J. W. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In *PCR Protocols—A Guide to Methods and Applications*. Academic Press.

Zak, B. (1976). Pure culture synthesis of Pacific Madrone ectendomycorrhizae. *Mycologia*, *68*, 362–369.

3.7. Figures and Tables

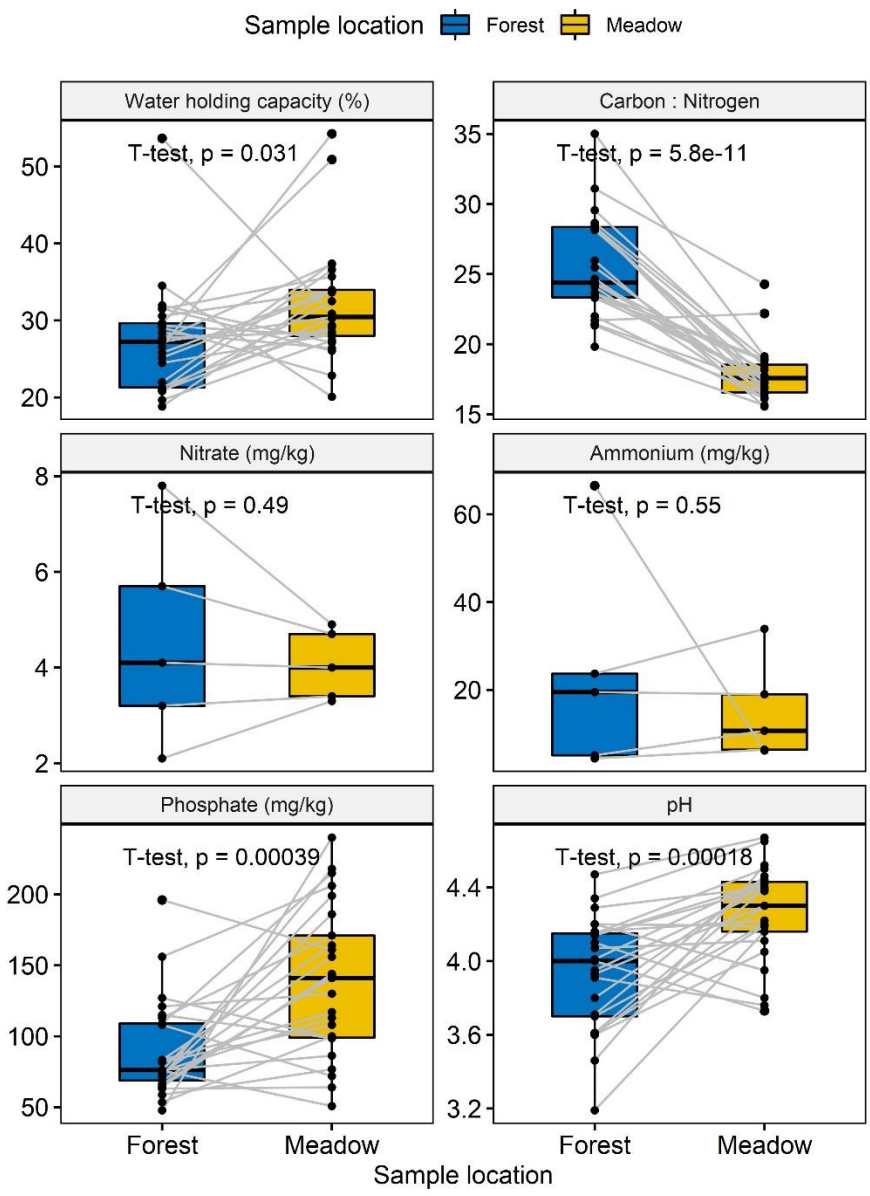

Figure 3.1. Chemical comparisons of soils sampled from the rhizospheres of mature *Abies lasiocarpa* growing in forest patches and subalpine meadows. Panel titles represent the label and units of y-axes and p-values refer to paired t-tests for each soil variable.

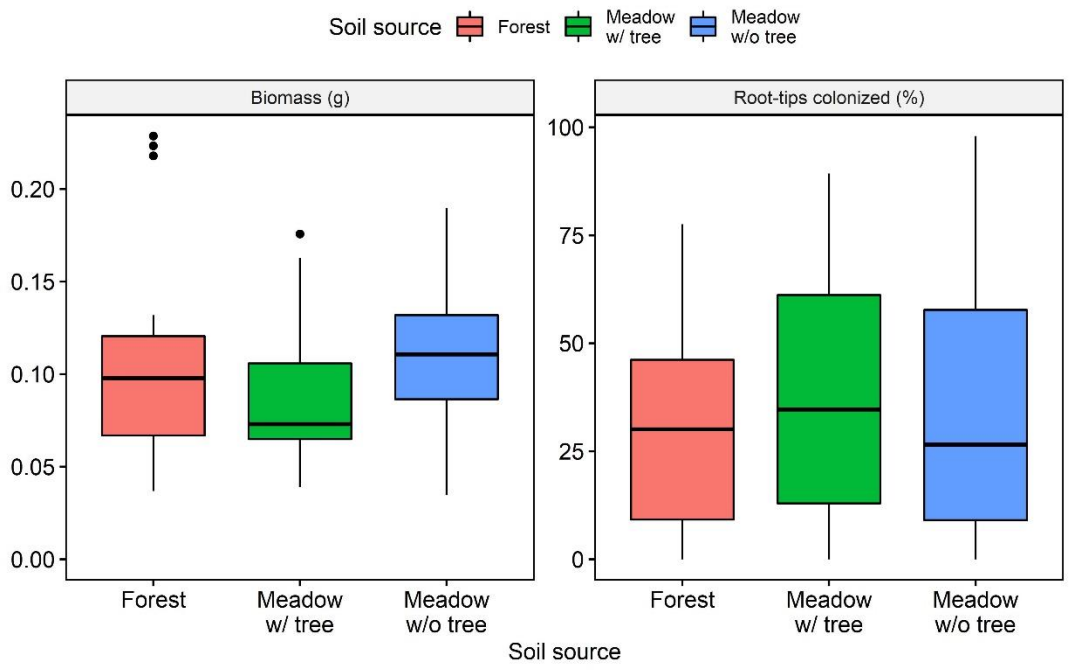

Figure 3.2. Comparison of biomass and extent of mycorrhizal colonization between *Abies lasiocarpa* seedlings grown in live soil sampled from different sources. Sources included the rhizospheres of mature conspecifics growing in forest patches (Forest) and subalpine meadows (Meadow w/ tree), and from an open meadow area with no nearby conspecifics (Meadow w/o tree). Panel titles represent the label and units of y-axes.

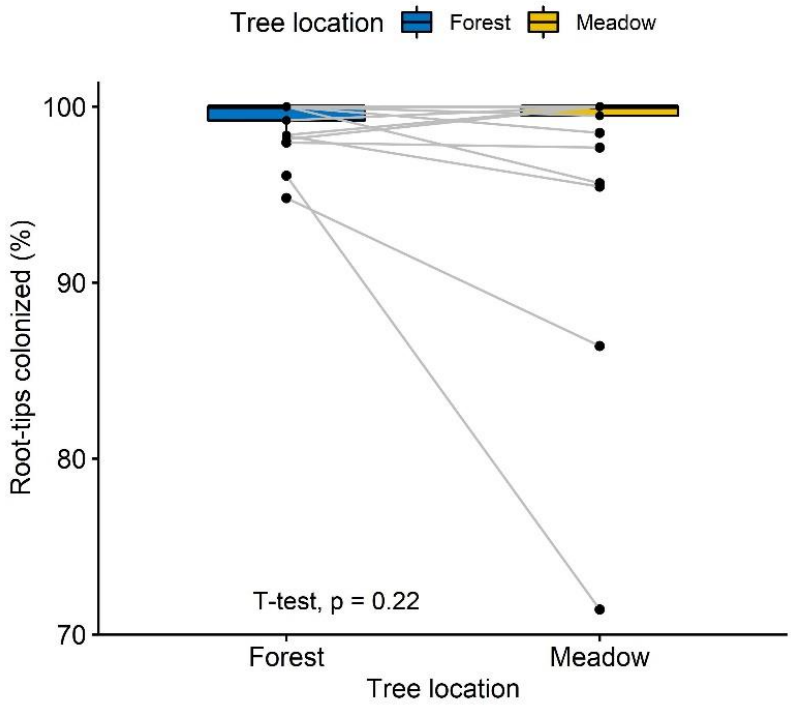

Figure 3.3. Extent of ectomycorrhizal colonization on roots of mature *Abies lasiocarpa* growing in forest patches and subalpine meadows. P-value represents a paired t-test.

Table 3.1. Dates and purposes of soil sampling events.

Sampling date	Measurement/use
24-25 September 2018	Water holding capacity, plant-available phosphorus, and C : N ratio
1 August 2019	Mature tree ectomycorrhizal colonization extent
2-4 September 2019	Plant-available nitrogen and pH
16 September 2019	Ectomycorrhizal inoculum for bioassay

Table 3.2. List of fungal taxa identified on roots of bioassay seedlings grown in different soil types based on matching of ITS sequences using NCBI BLAST (percent sequence identity ≥ 97%). Numbers in parentheses following taxonomic names indicate the number of distinct seedlings on which the taxon was found. For some seedlings grown in each soil type there was either no apparent mycorrhizal colonization or DNA extraction failed, which explains why the total number of seedlings referenced is lower than the 18 non-control seedlings per soil type.

Forest	**Meadow w/ tree**	**Meadow w/o tree**
Cenococcum geophilum (3)	*Wilcoxina mikolae* (9)	*Wilcoxina mikolae* (7)
Wilcoxina mikolae (2)	*Cenococcum geophilum* (2)	*Phialocephala fortinii* (1)
Amanita muscaria (1)	*Helotiales sp.* (2)	*Helotiales sp.* (1)
Hyaloscypha variabilis (1)		
Helotiaceae sp. (1)		
Helotiales sp. (1)		