A Machine Learning Approach to Tree Growth Modeling

Neal Maker and John Foppert

January 2021

1. Introduction

Tree growth modeling is important for forest managers, who use growth predictions to evaluate management approaches and establish forest valuations, but it is difficult in mixed-species and multi-aged forests where growth depends on particularly complex ecological relationships. For example, recent research in a number of different forest types has demonstrated that the species identity of neighboring trees can influence growth, and that the effects of species mixing depend on the species involved (Canham et al., 2004; Pretzsch, 2017 [chap 6 in mixed spp book]; Fien et al., 2019). In many cases, trees in mixed-species stands have higher potential growth rates than those in mono-specific stands because of positive interactions between species (Pretzsch and Zenner, 2017). Very few growth models account for species mixing, though, preventing managers from properly accounting for mixing effects in forest valuations and inhibiting efforts to take advantage of mixing effects through silvicultural interventions. Stand structural diversity is similarly absent from most operational growth models, but has been shown to relate to tree growth in complicated ways (Ali, 2019). Measures of structural diversity are positively correlated to growth in some settings (Danescu et al., 2016; Lei et al. 2009) and negatively correlated in others (Edgar and Burk, 2001; Liang et al., 2005), depending on the particular species and their development. In forests with many species growing in various arrangements these interactions become very hard to sort out, and they have not been studied at all in most forest types (O'Hara, 2014; Pretzsch et al. 2017).

One approach to modeling this complexity has been to shift analytical attention from stand- to neighborhood-dynamics (Canham and Uriarte, 2006). This is the approach taken by gap models, which conceptualize a forest stand as a collection of distinct patches of land (called 'gaps', 'neighborhoods' or 'plots') that

are typically 0.01 to 0.1 hectares in size and are assumed to be developing independently from one-another (Bugmann, 2001). Gap models' emphasis on neighborhood-scale dynamics grew out of ecological research demonstrating that plant-plant interactions are primarily local, neighbors' effects attenuate with distance, and beyond a certain distance plants have no detectable effect on one another (Stoll and Weiner, 2000). To simplify analyses and data collection, gap models do not account for the specific locations of trees within each neighborhood (they are distance-independent). The gap model framework therefore offers a compromise between distance-dependent models, which are fine grained but difficult to apply, and stand-scale models, which are easy to apply but do not account for local effects directly (Bugmann, 2001). Historically, gap models have been primarily used for modeling long-term system-level dynamics (e.g., Forsyth et al., 2015) and the majority of research attention has been directed toward representing regeneration, recruitment, and mortality based on an understanding of ecological processes (Weiskittel et al., 2011). These 'mechanistic' models rely on causal relationships and are well suited to studying ecosystem changes (Cuddington et al., 2013), but are often less accurate than empirical models in predicting overstory tree growth, more difficult to initialize, and incapable of capturing ecological processes that have not been well-studied (Taylor et al., 2009), limiting their application for practitioners. Many empirical forest growth models have been built at the stand scale for use in operational decision making (Weiskittel, 2011), which could be adapted to a gap model framework to address these concerns. They are generally parametric models constructed with traditional statistical techniques (ibid.). In this paper, we expand on that idea and explore the potential for integrating empirical machine learning models into the gap model framework instead.

Machine learning (ML) is a branch of artificial intelligence that uses data to "learn" algorithms for

predicting responses, making no assumptions about the processes that generated the data. This is in contrast to the traditional statistical approach to modeling (referred to here as the "a priori" approach), in which the data are assumed to have come from a stochastic model and previous research guides the selection of a model form and the predictors to be used (Breiman, 2001). Breiman (2001) points out three advantages of ML over a priori modeling, which he refers to as the three important lessons of ML: *First*, ML recognizes that many different models can do a good job explaining the same phenomenon and takes an experimental approach to model selection; *second*, ML does not shy away from complex model formulations when they can improve predictions; and *third,* ML accommodates high-dimension data, seeking relationships among many interrelated variables. For these reasons, ML has been successful in modeling complex processes with interrelated predictors, and in situations where extensive or high-dimension data are available.

Growth modeling in mixed-species, multi-aged forests is just such an application. Growth is driven by complex relationships between various factors and, in many places, large amounts of forest inventory data have been collected over extended time periods (Kershaw et al., 2017). Forest ecologists have begun using ML more, especially for predicting nonlinear ecosystem behaviors (Liu et al., 2018), but its application to forest growth modeling has mostly been limited to tree mortality (e.g., Fan et al., 2006; Guan and Gertner, 1991; Hasenauer and Merkl, 1999; Kershaw et al., 2017; King et al., 2000; Sironen et al. 2010), which has been especially hard to predict using a priori models.

We illustrate the ML approach by constructing a gap-scale, tree diameter increment model, using data from the ecologically complex United States Northern Forest region. We compare the ML model to a second, a priori model that we parameterized to the same dataset. We also demonstrate some of the model-agnostic interpretation tools that exist and how they can be applied to complex forest growth models to

visualize relationships between predictors and outcomes, and discuss the potential for ML in studying tree growth.

2. Data

Data came from the US Forest Service's Forest Inventory and Analysis (FIA) program (https://apps.fs.usda.gov/fia/datamart/datamart.html, downloaded December 6, 2020) from remeasured plots in the 36-county Northern Forest region. This region includes the New York Adirondacks and northern Vermont, New Hampshire, and Maine (Figure 1). We used data from the current national standard plot design, which was adopted in the mid-1990s (Woudenberg et al., 2010). These plots are systematically located and evenly distributed across the region. Each plot is made up of four 168 m^2 subplots with a 13.5 m^2 microplot inside each subplot. Trees greater than 12.7 cm diameter at breast height (DBH) are measured on subplots, and trees between 2.54 and 12.7 cm DBH are measured on microplots.

FIA data relate to numerous tree-, plot-, and stand-level attributes, and are collected through ground sampling and remote sensing. Our objective was to build a model for operational decision making, so we focused on attributes that can be measured reasonably easily by practitioners. They include DBH, species, compacted crown ratio (CR), crown class, tree class (a measure of general quality), forest type, stocking level, landscape (physiographic) position, site class, slope, aspect, elevation, latitude, and longitude. We also used existing data to calculate a number of new attributes, including subplot basal area (BA), tree overtopping basal area (BAL), and species-specific subplot BA and tree BAL. To keep the data and resulting model relevant to practitioners, we grouped tree species into species groups and FIA forest types into more general forest types. Our groupings were designed to match common inventory protocols that we have

observed. Finally, we removed trees from the analysis that died during the remeasurement period, and removed all plots with missing data. The final dataset includes 399,358 observations from 11,286 different plots. Remeasurement intervals on the plots ranged from 2.9 to 7.8 years and averaged 5.1 years.

Data from 20% of the plots was set aside for testing final models, to account for over-fitting and allow a fair comparison between the machine learning approach and the a priori approach. Data were split based on plots to make sure that the testing data are truly independent. The data from the remaining 80% of plots were used for exploratory analysis, to compare different modeling methodologies, and to train the final ML and a priori models.

All of the data acquisition and processing, and the analysis that follows, was done using the programming software R, version 4.0.1 (R Core Team, 2020). Much of the general data manipulation and visualization was carried out with R Tidyverse packages (Wickham et al., 2019); and models were trained, tuned and compared using Kuhn's (2020) "caret" framework. Detailed methods are available at https://github.com/nealmaker/tree-growth-nf.

3. Study Area

The Northern Forest is located between the temperate forests of the eastern US and the boreal forests of Canada, and contains characteristics of both biomes, giving it a

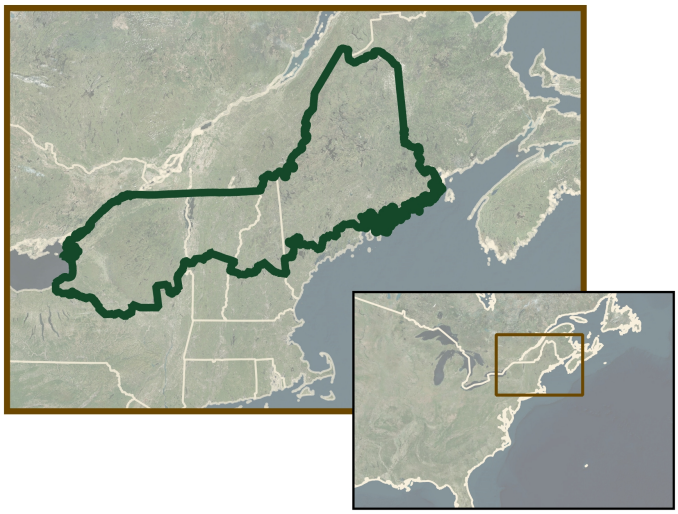

Fig 1. Northern Forest region, depicted with dark outline.

diverse mixture of tree species and site conditions. Trees in the training data belong to 28 different species groups and are located in 17 different forest types and 14 different landscape positions. The most common species groups (by number of stems) are fir, spruce, soft maple, hard maple, cedar, beech, yellow birch, paper birch, hemlock and white pine. Plots' elevations range from 0 to 1,292 meters above sea level.

The region is well suited to studying the complexities of tree growth because its forests tend to be complex. Most are naturally regenerated, second-growth forests with varied management histories. In the training data, some 68 % of plots contain three or more species growing together, while the spread of within-plot tree diameters averages 35 cm.

4. Methods

4.1 A Priori Approach

To illustrate the a priori approach to growth modeling, we built a nonlinear least squares model of individual tree diameter increment, based on a very similar model that Weiskittel et al. (2016) built for the New York Adirondacks. Our model used the same form as theirs:

$$ADI = \exp\left(b_1 + b_2 \cdot \ln\left(DBH\right) + b_3 \cdot DBH + b_4 \cdot BAL + b_5 \cdot \sqrt{BA} + b_6 \cdot SC\right)$$

where ADI is the annual diameter increment ($cm \cdot yr^{-1}$), DBH is diameter at breast height (cm), BAL is overtopping basal area ($m^2 \cdot ha^{-1}$), BA is plot basal area ($m^2 \cdot ha^{-1}$), and SC is site class. These predictors have been well studied in the Northeastern US in relation to diameter increment (Teck and Hilt, 1991; Westfall, 2006) and their use in predicting diameter increment has a strong conceptual basis (Weiskittel et al., 2011).

The equation was fit to the training data for each species group to obtain species-specific coefficients.

One difference between our model and that of Weiskittel et al. is that we used site productivity class as a measure of site class, which the FIA program calculates. It is an ordinal categorical variable based on the potential wood volume growth on a site. Weiskittel et al. used climate site index, which is a climate-based estimate of site index. Another difference is that Weiskittel et al. used nonlinear mixed-effects modeling (NLME). Their data came from continuous forest inventory plots in five forests, and they treated forests as a random effect. We experimented with using NLME and treating plots as random, but found no improvement over nonlinear least squares modeling. Both approaches have been used successfully to model individual tree growth in the Northeast (Kuehne et al., 2020).

We used McDill and Amateis' (1993) nonlinear "two-step" interpolation method to account for the fact that different plots in the training data had different remeasurement intervals. The method assumes that the diameter growth rate over the period varies according to the model form. Nonlinear interpolation has been shown to be more accurate than assuming a linear growth rate over the remeasurement period (McDill and Amateis, 1993; Cao, 2000), or using the remeasurement interval as a predictor (Juma et al.,2014).

4.2 Machine Learning Approach

The ML approach depends on extensive data to determine the most effective modeling methodology; and allows for the assessment of a large number of predictors, even if they are poorly understood to begin with. An organized, experimental framework is key to the approach (Kuhn and Johnson, 2013). We break the approach down into four main elements following Kuhn and Johnson (2013) and Boehmke and Greenwell (2020): (1) pre-processing, (2) resampling, (3) spot-checking, and (4) final

tuning. To our knowledge, no model-agnostic, nonlinear interpolation methods exist for annualizing remeasurement data, so we assumed linear growth through the remeasurement period and trained ML models on average annual DBH growth rates.

4.2.1 Pre-processing

Different modeling methodologies vary in their handling of categorical data and make different assumptions about variables' distributions and relationships. Pre-processing is used to massage data into a form that can be widely used across methods. It can also increase model accuracy, particularly with methods that are sensitive to skewed data or to predictors on very different scales (Kuhn and Johnson, 2003). We used a number of pre-processing steps on predictors. Nominal categorical predictors were dummy-encoded, to turn them into binary predictors and allow for their use in various regression models; numeric predictors were normalized using Yeo-Johnson transformations (Yeo and Johnson, 2000), centered, and scaled, to standardize them and distribute them normally; and highly correlated predictors were filtered to avoid near-perfect collinearity and prevent problems with linear-based methods.

4.2.2 Resampling

The a priori approach generally prevents model over-fitting by using simplified, predetermined model forms that cannot be bent out of shape to fit the noise in the training data. Many ML methods are more flexible, and over-fitting must be deliberately prevented. This is generally done using resampling and model tuning. Tuning parameters in the model determine its flexibility, and resampling techniques are used to estimate the true prediction error. Numerous tunings are tested, and the one that minimizes the

estimated prediction error is used. The optimal tuning will be flexible enough to capture complex signals in the data, but not so flexible as to include excess noise (Hastie et.al., 2017). We used ten-fold cross validation for resampling, which breaks the training data into ten random subsets, or *folds*, and trains a model multiple times, holding out each fold in turn to use as independent testing data. Prediction errors are calculated for each fold and averaged to estimate the true prediction error. The pre-processing steps described in the previous section were applied independently to each training in the cross validation, to prevent information from the reserved fold from biasing the estimated prediction error. This was done using using the "recipes" framework developed by Kuhn and Wickham (2020).

4.2.3 Spot-Checking

A wide variety of modeling methods were compared using cross validation. We were most concerned with accuracy, which we measured using root-mean-squared error (RMSE), but we also compared prediction speed and model object size (megabytes) to account for how easily models could be incorporated into growth simulators. Comparisons can be based on whichever metrics are most meaningful for any given problem. Spot-checking was designed to quickly identify the most promising types of modeling algorithms, and we used a random subset of the training data and parallel processing to speed the process. We performed minimal tuning of each model using default methods.

Eighteen different algorithms were successfully spot-checked. They were chosen to represent a wide variety of methodologies and included linear regression (R Core Team, 2020), ridge regression (Zou and Hastie, 2020), elasticnet (Zou and Hastie, 2020), penalized linear regression (Goeman et al., 2018), lasso-regularized generalized linear modeling (Friedman et. al., 2010), quantile regression with lasso penalty

(Sherwood and Maidman, 2020), multivariate adaptive regression splines (Milborrow, 2020), bayesian ridge regression (Gramacy et al., 2019), k-nearest neighbor (Schliep and Hechenbichler, 2016), one layer (Venables and Ripley, 2002) and multilayer perceptitron (Bergmeir and Benítez, 2012) neural networks, linear and radial support vector machines (Karatzoglou et al., 2004), a regression tree (Therneau and Atkinson, 2019), bagged trees (Peters and Hothorn, 2019), a Random Forest (Wright and Ziegler, 2017), a gradient boosting machine (Greenwell et al., 2019), and a model tree (Kuhn and Quinlan, 2020).

The results demonstrated that, for DBH increment in the Northern Forest, nonlinear methods like model trees, random forests, gradient boosting machines, and support vector machines tend to be the most accurate. Regularized parametric models are not as accurate, but tend to be smaller and faster at making predictions. We chose a gradient boosting machine as the best compromise in this case, because it offered high accuracy with relatively fast prediction speeds, and was reasonably small compared to many of the other nonlinear methods.

4.2.4 Final Model Tuning

Having chosen the best algorithm for the application, we retrained the model using all of the training data and a more organized tuning process. Gradient boosting machines are particularly sensitive to their tuning parameters and difficult to train efficiently, and we used a process described by Boehmke and Greenwell (2020, section 12.3.3). Five-fold cross validation was used instead of ten-fold to speed the process, because it was very computationally intensive. Our gradient boosting machine is an ensemble of weighted regression trees, and the tuning parameters that determine its complexity and flexibility are the shrinkage rate, the number of trees in the ensemble, the minimum terminal node size of each tree, and the

maximum depth of each tree (Greenwell et al., 2019). The final tuned model has a shrinkage rate of 0.5 and uses 14,000 trees, with each tree having a minimum terminal node size of 15 observations and a maximum tree depth of two branches. See Hastie et al. (2017, pp. 337-388) for a more detailed description of gradient boosting trees and their tuning parameters.

4.3 Model Assessment

The a priori and ML models were compared based on their ability to accurately predict DBH growth in the independent testing data. The models predict annualized growth increments, while growth was actually observed over multi-year remeasurement intervals. To predict whole-interval growth (ΔDBH) comparable to observed growth, we applied the models iteratively to each observation, predicting one year's growth at a time and updating the DBH, BA and BAL before predicting the next year's growth. We assumed a linear change in BA and BAL over each remeasurement period (after Weiskittel et al., 2016). Because remeasurement intervals varied between plots, the ΔDBH values we report are not directly comparable to one another. Observed and predicted ΔDBH are on the same scale for each observation, though, allowing for a fair comparison between the a priori and ML models. Overall accuracy was assessed using RMSE and mean absolute error (MAE).

We also used several model-agnostic interpretation tools to explore the structure of the ML model. Nonparametric models like the gradient boosting machine can be very complicated, and instead of trying to fully map the model space, we chose a couple of examples to illustrate the interpretive process. To start, we calculated permutation feature importance (Fisher et al., 2019) for the predictors. This technique involves shuffling the values for a given feature (predictor) on the observations in a dataset, and observing the

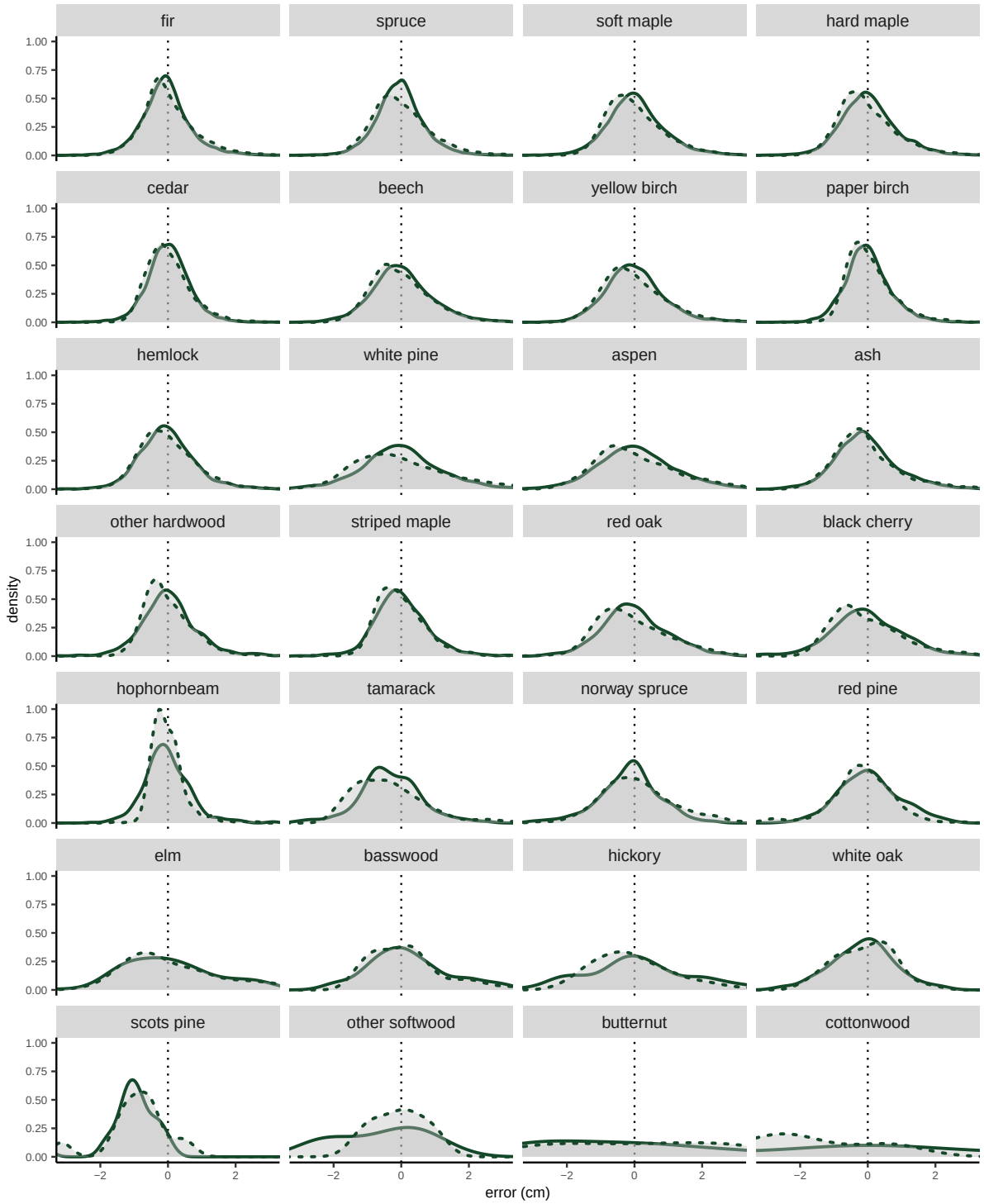

Fig 2. Prediction error distributions (observed – predicted) for individual species groups, for the gradient boosting machine (solid line) and nonlinear least squares model (dashed line). Distributions are drawn using kernal density estimation. Positive errors are under-predictions and negative errors are over-predictions.

resulting changes in accuracy. It accounts for predictors' main and higher order effects, and shows which predictors are most important overall. Test data was used to compute feature importance (so that it measures features' actual importance in making new predictions rather than their importance in training the model) and RMSE was used to assess accuracy. Then we used Friedman's H-statistic to search for potential interactive effects (Friedman and Popescu, 2008) and visualized several of them with partial dependence plots (Friedman, 2001). The H-statistic can be used to calculate the overall interaction strength for a predictor, across all other predictors, and to calculate a specific interaction strength between two predictors. We calculated overall strengths for all of the predictors and used the results to guide our calculation of specific strengths, teasing out some of the higher-order effects in the model. The H-statistic and partial dependence plots are based on the same underlying theory as permutation feature importance, and use permutation of observations in the data to isolate the effects of one or more predictors (Molnar, 2020). We implemented all three techniques using Molnar et al.'s (2018) "iml" package for R.

Results

The mean observed ΔDBH was 1.27 cm, for remeasurement intervals that averaged 5.1 years. The distribution of observed ΔDBH was skewed right, with 90% of observations falling between 0 and 3 cm. Results show that the gradient boosting machine built with the ML approach achieved a marked improvement in accuracy over the a priori, nonlinear least squares model, reducing the RMSE by 6.5% and the MAE by 7.9% (Table 1). The improvement clearly came at the cost of speed and size, though, with the ML model taking 27 times as long to make predictions and using almost 44,000 times the space in memory.

Metric	NLS Model	ML Model
RMSE	0.364	0.341
nRMSE	0.730	0.683
MAE	0.269	0.248
Speed (s)	0.07	1.97
Size (MB)	0.003	131

Table 1. Performance metrics for the a priori nonlinear least squares model (NLS) and the ML gradient boosting machine (ML). nRMSE is the normalized RMSE, expressed as a proportion of the mean observed response, and speed is the time to make 1,000 predictions on our computer, and is relative.

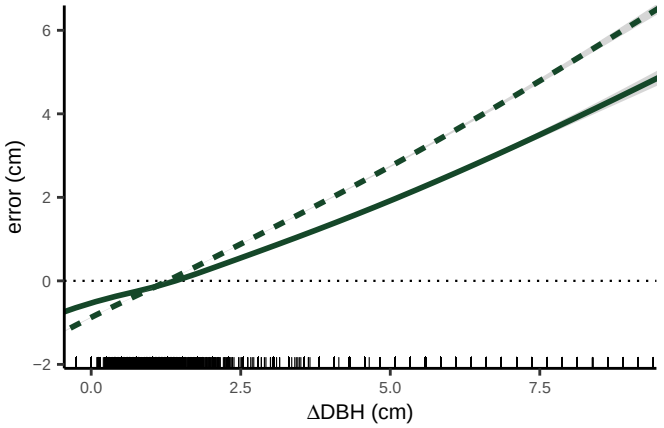

Fig 3. Prediction error trends (observed – predicted) based on DBH, for the gradient boosting machine (solid line) and nonlinear least squares model (dashed line). Trends are drawn with generalized additive models

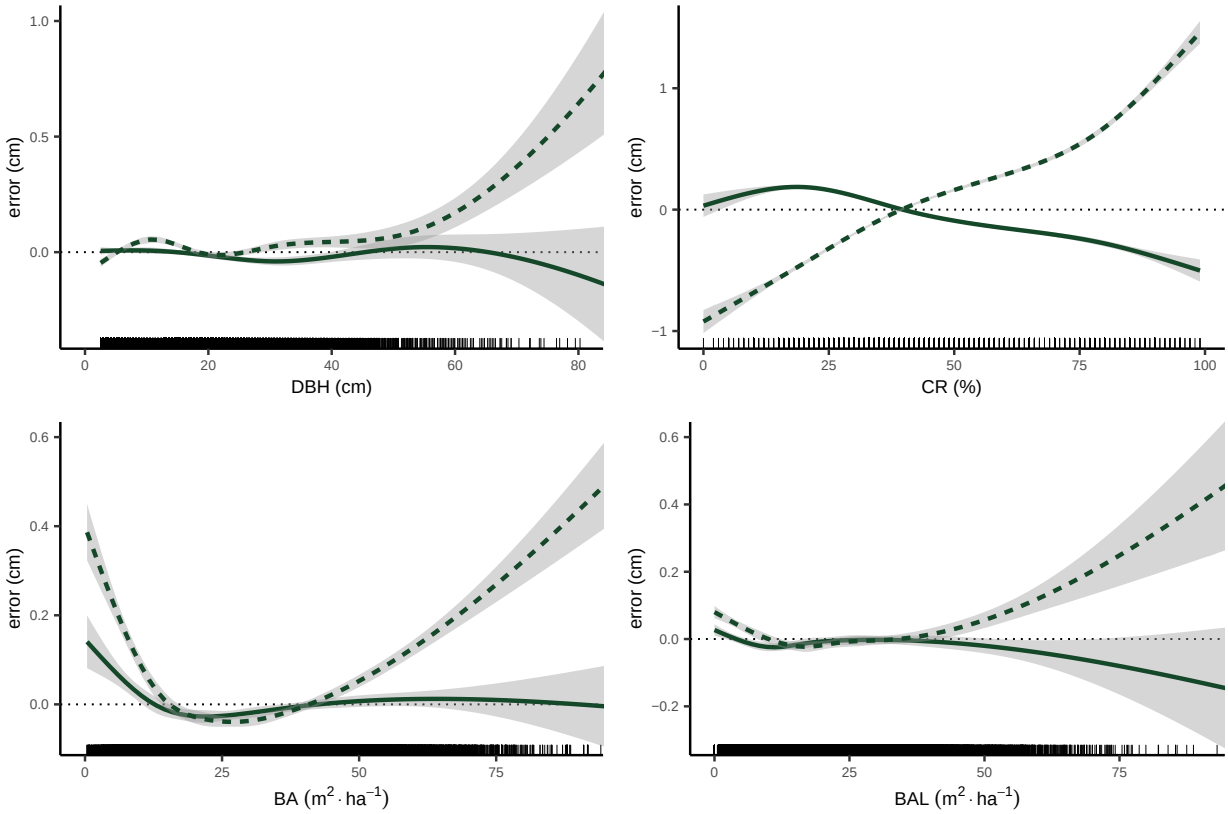

Fig 4. Prediction error trends (observed – predicted) based on four predictors, for the gradient boosting machine (solid line) and nonlinear least squares model (dashed line). Trends are drawn with generalized additive models.

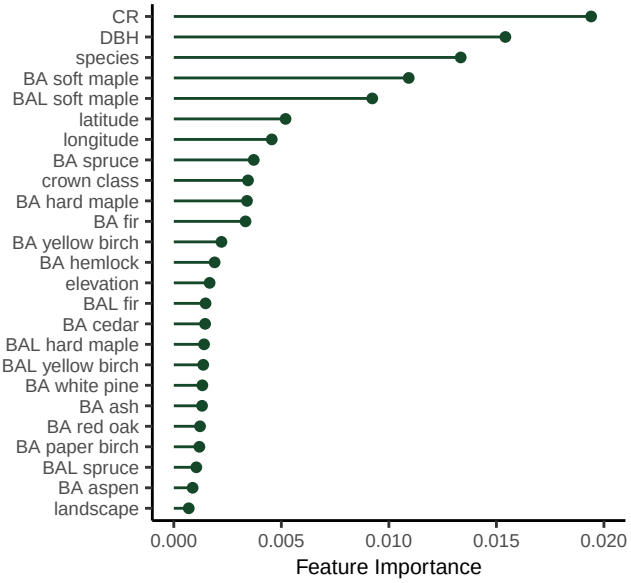

Fig 5. Permutation feature importance of the 25 most important predictors. BA and BAL are decomposed by species.

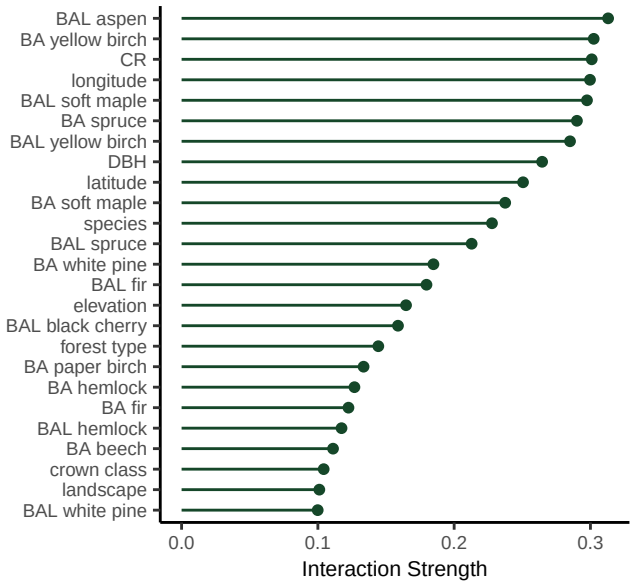

Fig 6. Interaction strengths of the 25 predictors demonstrating the most interaction, calculated using Friedman's H-statistic. BA and BAL are decomposed by species.

Both models are relatively unbiased overall, with biases of 0.003 cm for the ML model and -0.005 cm for the a priori model (observed - predicted). Both are relatively unbiased for common species too (Figure 2). The ML model is less biased than the a priori model in predicting the growth of especially slow- and fast-growing trees (Figure 3) and in predicting growth in trees with less common attributes and in less common settings (Figure 4).

Basal area is the most important predictor in the ML model (with a feature importance score of 0.035), followed by CR (0.019), BAL (0.019), DBH (0.015), and species (0.013). Figure 5 shows the relative importance of the 25 most important predictors, with species-specific BAs and BALs listed separately.

The predictors with the highest overall interaction strengths in the ML model are shown in Figure 6. By also calculating two-way interaction strengths between predictors with

high overall strengths, we uncovered several important interactions that influence the model. The most important is a three-way interaction between species, DBH, and CR. Three-way interactions are inherently difficult to visualize, and we illustrate it in Figure 7 for only three of the 28 species groups. Generally, we see that the effect of CR is amplified at larger DBH, the effect of DBH is amplified at larger CR, and both the CR and DBH effects are larger in some species (e.g., white pine) than in others (e.g., hophornbeam). A second interesting interaction occurs between fir BA and species (Figure 8). For most species, the model predicts reduced DBH growth on plots with higher fir BA, but fir BA does not affect predicted growth for other species like

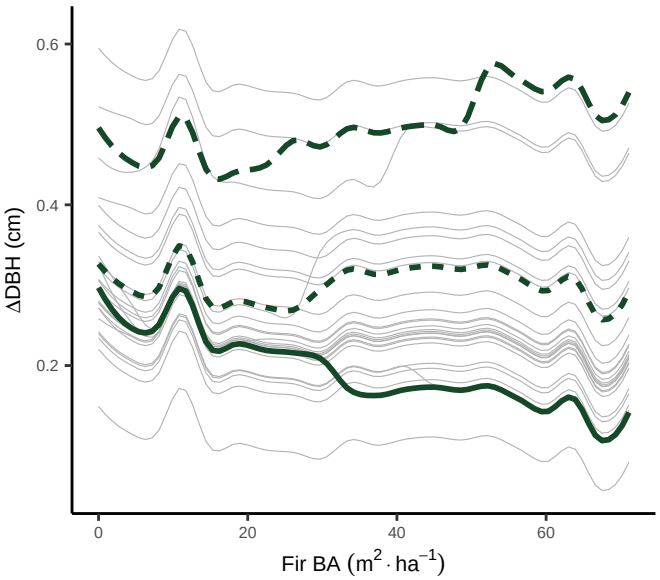

Fig 8. Relationship between fir BA and predicted ΔDBH for individual species, using partial dependence plotting. Three species have been highlighted: aspen (dashed line on top) demonstrates increased growth at higher fir BA, hemlock (dotted line in middle) demonstrates growth poorly correlated to fir BA, and yellow birch (solid line on bottom) demonstrates reduced growth at higher fir BA.

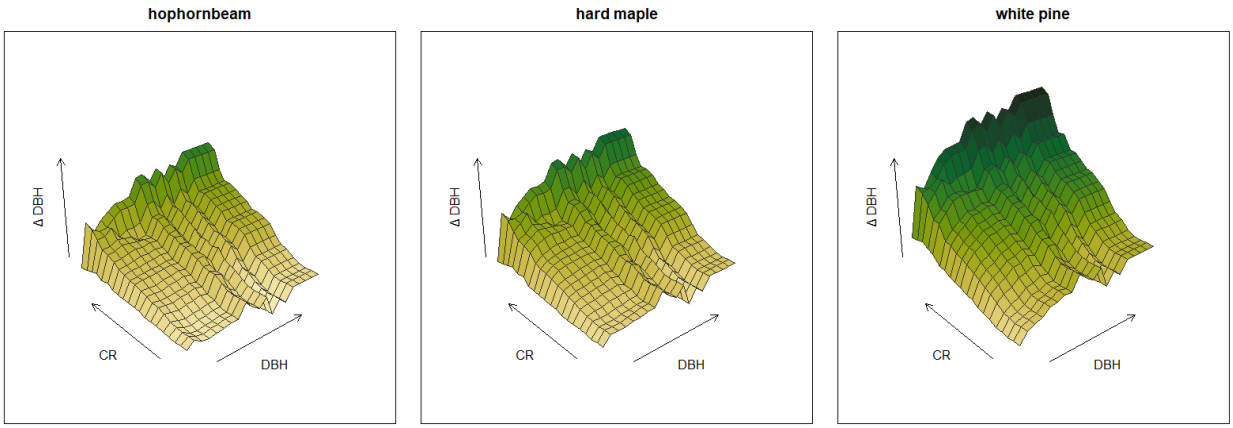

Fig 7. Relationship between DBH, CR, and predicted ΔDBH for three individual species groups, using partial dependence plotting. The shape of the response curve for each predictor changes depending on the other predictors, indicating that the predictors interact with one another.

hemlock. Aspen actually experiences increased predicted growth rates at higher fir BA. There may be many more interactions driving the ML model as well, which we did not explore. Interaction strengths suggest interactions related to latitude and longitude, and to a number of different species-specific BAs and BALs.

Discussion

Our ML and a priori models perform similarly overall and for average trees in common situations, but the ML model is more accurate for trees in abnormal situations, such as for larger diameter trees, trees with above- or below-average CR, and trees growing with above-average competition. Part of that advantage likely comes from the ML model's use of numerous predictors. While the a priori approach singles out the most important predictors, the ML approach also makes use of less important predictors whose simple effects might add to the predictive ability. Crown ratio is quite important to the ML model's structure, for example, and is absent from the a priori model. Previous a priori models of diameter growth in the region rejected CR in favor of other measures of competition (Teck and Hilt, 1991; Kiernan et al., 2008; Weiskittel et al., 2016) and we also left it out of our a priori model in favor of BA (as a measure of two-sided competition) and BAL (as a measure of one-sided competition). The ML approach retained all of these measures of competition, along with crown class, stocking, and numerous species-specific BAs and BALs, and exploited any that contributed to the specific algorithm in use. Latitude and longitude are also reasonably important to the ML model, pointing to a geographic aspect of diameter growth that the a priori approach ignored.

The ML model also accounts for some complex interactions that the a priori model does not capture. We illustrated the three-way interaction we found between species, DBH, and CR; and the uneven

298 effect of fir BA on different species as examples. Interaction strengths pointed to several more potential
299 interactions, and there could be many more among less important predictors, which would contribute little
300 to overall accuracy, but could be important in specific situations. All three of Breiman's (2001) "lessons"
301 come into play in uncovering these interactions: *First,* the ML approach did not limit model selection to
302 match a preconceived conception of growth dynamics, and was able to find a particularly good model
303 algorithm that had not been tried in this context before; *second,* the ML approach did not impose limitations
304 on model complexity, and employed a complicated model more capable of representing interactions, and
305 *third,* the ML approach embraced the high dimensional data, which contained information about interactions
306 that would have been missed otherwise.

307 The complex, flexible models that ML can produce are especially good at uncovering the subtleties
308 of neighborhood-scale dynamics, especially when they are trained on fixed area plot data (like FIA data) that
309 came from the direct measurement of tree neighborhoods. In this sense, they are a natural fit for the
310 neighborhood-oriented gap model framework. Gap-based simulators are already well developed (Bugmann,
311 2001), but many are focused toward modeling regeneration, recruitment, and mortality processes rather
312 than overstory tree growth (Weiskittel et al., 2011), resulting in some cases in significant biases in growth
313 projections relative to other modeling approaches (e.g., Yaussy, 2000). Unbiased, emperical ML models (of
314 diameter growth, height growth, and crown growth, for example) could easily be used in place of existing
315 submodels to improve growth predictions and make gap-based simulators more suitable for practitioners.

316 One critique of empirical models like these has been their unsuitability for making predictions in
317 novel scenarios. Their structure blindly mimics the data used to train them, and there is no guarantee that
318 the relationships present in the data will hold in the future (Cuddington et al. 2013). Process-based models

account for the underlying processes that drive growth, and are better suited to long-term predictions; but they depend on complex, poorly understood causal relationships that still need to be comprehensively studied (Ali, 2019; Pretzsch et al., 2017 [modeling chapter]). Experimental research into these relationships often depends on drawn out spatial-temporal studies (Pretzsch, 2017 [chap 6 in mixed spp book]) and will take many years. For forest managers, short- and medium-term accuracy is often most important, and current process-based models cannot compare with empirical models (Taylor et al. 2009). This may change as our understanding of causal relationships grows.

In most cases ML is blind to causality, but it can be employed to find potentially causal relationships, to guide experimental research and speed the development of process-based models. Existing forest inventory data contains a wealth of information for ML techniques to mine; and the rapidly expanding new field of interpretable ML is devoted to studying ML models to understand the relationships they glean from data (Molnar, 2020). Variable importance, interaction scores, and partial dependence plots are a few of the better established techniques, but there are many more, and they are growing more sophisticated as the field matures. Models that used to be considered black boxes are now becoming powerful microscopes. For the time being, and thanks to a growing body of forest inventory data, ML could be useful for studying tree growth in addition to predicting tree growth.

This example shows that the ML approach has potential to increase the accuracy of tree growth models for complex forests, and can further refine and compliment the individual-tree, gap model framework that is already in wide use. It also demonstrates the potential for ML to exploit extensive forest inventory data to study ecological relationships pertinent to tree growth. These benefits will only become more apparent as the fields of ML and interpretive ML grow, and as more forest inventory data becomes

available.