Nonparametric individual-tree growth models for the US Northern Forest

Neal Maker & John Foppert

April 10, 2020

Introduction

Numerous tree growth models have been developed for the forests of the Northeast, which vary in their methodological approaches and in their applicability. As focus has shifted from plantation- to natural forest-management and as computing power has increased, there has been a movement from stand-level approaches (which use standwide metrics and are better suited to even-aged, monospecific stands) to individual-tree approaches (which consider the attributes of individual trees and are better suited to structurally complex, multi-species stands) (Peng 2000). Simultaneously, interest in process-based models has grown because of their superiority making long-term predictions–especially in the face of climate change (Cuddington et al. 2013). Process-based models mimic known ecological processes, so they can be applied to novel scenarios as long as those underlying processes don't change. Emperical models, which ignore the underlying ecological processes, are only applicable to scenarios that are represented in the data used to build them. More purely process-based models are therefore well suited to research, but not as well suited to forest management planning because they are more complex, often depend on difficult to measure ecological parameters, and are generally not as accurate over shorter time-scales (Peng 2000).

Modern operational models usually strike a balance by using a more emperical approach that is still structured around an ecologically informed conception of growth dynamics. They are generally parametric models, with predetermined structures to ensure that they are somewhat mechanistically realistic, but with model coefficients determined emperically. For example, A. Weiskittel et al. (2016) used nonlinear mixed effects modeling to predict total height, bole height, diameter increment, height increment, and mortality in the Adirondacks, with different predictors and equation forms for individual submodels based on previous knowledge of the relevant relationships. Westfall (2006) and Westfall and Laustsen (2006) used very similar mixed effects models to predict diameter increment in the Northeast and height in Maine, respectively. And Teck and Hilt (1991) used a somewhat more mechanistic approach for their Northeastern diameter growth model: first predicting the maximum potential growth rate based on species, diameter, and site index, then modifying that prediction downward based on overtopping basal area. All of these models use a handful of predictors whose effects had been studied previously.

Meanwhile, the rise of big data in the last several decades has led to the development of sophisticated nonparametric models, which do not require a priori specification of model structures and do not rely on any assumptions about populations' probability distributions. Nonparametric models are purely emperical and excel in situations where processes are poorly understood. They have not been widely adopted by foresters, and have several obvious disatvantages to conventional parametric models. First, they cannot be applied to novel situations and are especially bad at making long-term predictions in a changing climate. Second, they fail to incorporate the knowledge gained from past studies, so using them is kind of like starting over instead of building on the decades of research that is available. And thirdly, nonparametric models tend to be much more complex than parametric models: harder to interpret, more computationally intensive, and less accessible to forest managers.

At the same time, these characteristics give nonparametric models some real advantages, especially in the ecologically complex Northern Forest. While parametric models must stick to a handful of well-researched predictors that can be accurately specified, nonparametric models can account for poorly-understood predictors; and can capture interactions between multiple predictors–which are more likely to exist in complex ecosystems–even if those interactions were previously unknown. Also, many nonparametric models are capable of handling large amounts of data and large numbers of predictors simultaneously, allowing a more holistic approach to growth modeling. Even if these advantages don't appreciably increase models' overall accuracy, they make nonparametric models more nuanced and likely decrease bias at finer scales (capturing the subtle differnces between trees in different situations). And big strides have been made recently in interpreting nonparametric models, so that the complex relationships they uncover can be seen and studied (Goldstein et al. 2015; Molnar 2020).

This paper describes our development of nonparametric diameter increment, height, height increment, survival, and crown ratio change models for the US Northern Forest. Our principal goals in building the models were to (1) allow foresters to make short-term predictions that better account for ecological complexity, and can be used to plan irregular management regimes, and (2) study the growth effects of poorly understood factors and the interactions between factors, to add to our understanding of Northern Forest growth dynamics.

Data

We took advantage of the large, statistically rigerous Forest Inventory and Analysis (FIA) database that the US Forest Service maintains, which tracks numerous tree, stand, and site characteristics.

The model described here is built on a considerably larger sample than previous models, which was drawn from the US Northern Forest region (Maker 2019). The Northern Forest covers a fairly broad geographic area while still representing a coherent socio-ecological unit–in which trees can be expected to follow a similar set of patterns.

Each model was trained against more than 200,000 observations of individual trees, drawn from more than 10,000 remeasured inventory plots, using sixteen predictive variables.

Analysis

We chose tree-based, Random Forest algorithms for their ability to handle many predictors, their computational efficiency, and their easy integration of numeric and categorical predictors. A number of tools were used to interpret the models, to quantify the importance of individual predictors and to visualize the relationships between predictors and growth outcomes.

A random forest algorithm was chosen to train the model, for a number of reasons. first, random forests can handle a large number of predictors, unlike generative models (which would require many parameters and be subject to overfitting) or nearest neighbor-type algorithms (which suffer the "curse of dimensionality"). Second, they are computationally efficient, especially when being used for prediction. Because the model will be used by practitioners "in the field", this will be a major asset. Finally, random forest algorithms can account for the interactions between numeric and categorical predictors, like those between species and latitude, or between stand basal area and forest type. Interactions like these are not accounted for in most existing growth models, and they can help to describe some of the variation in diameter growth rates, increasing the accuracy.

Results

Discussion

Competition indices can still be obtained from conventional (non-spatial) inventory techniques, and in many cases they can be used to derive growth estimates with overall accuracies comparable to those derived from spatially-explicit competition indices (Kuehne, Weiskittel, and Waskiewicz 2019). Still, subtler growth effects at the individual tree scale are probably better accounted for by spatially explicit modeling, and these are particularly important to the management of complex irregular forests. Also, spatially explicit, process based models are much more capable when it comes to modeling novel scenarios, like tree growth in a changing climate, and the effects of new management regimes (Cuddington et al. 2013).

Literature Cited

Cuddington, K., M.-J. Fortin, L. R. Gerber, A. Hastings, A. Liebhold, M. OConnor, and C. Ray. 2013. "Process-Based Models Are Required to Manage Ecological Systems in a Changing World." doi:10.1890/ES12-00178.1.

Goldstein, Alex, Adam Kapelner, Justin Bleich, and Emil Pitkin. 2015. "Peeking Inside the Black Box: Visualizing Statistical Learning with Plots of Individual Conditional Expectation." *Journal of Computational and Graphical Statistics* 24 (1): 44–65. doi:10.1080/10618600.2014.907095.

Kuehne, Christian, Aaron R. Weiskittel, and Justin Waskiewicz. 2019. "Comparing Performance of Contrasting Distance-Independent and Distance-Dependent Competition Metrics in Predicting Individual Tree Diameter Increment and Survival Within Structurally-Heterogeneous, Mixed-Species Forests of Northeastern United States." *Forest Ecology and Management* 433 (February): 205–16. doi:10.1016/j.foreco.2018.11.002.

Maker, Neal. 2019. "Development of an FIA Dataset to Model Tree-Level Changes in the Northern Forest." https://github.com/nealmaker/fia-data-nf.

Molnar, Christoph. 2020. *Interpretable Machine Learning*. https://christophm.github.io/interpretable-ml-book/.

Peng, Changhui. 2000. "Growth and Yield Models for Uneven-Aged Stands: Past, Present and Future." *Forest Ecology and Management* 132: 259–79.

Teck, Richard M., and Donald E. Hilt. 1991. "Individual Tree-Diameter Growth Model for the Northeastern United States." *Res. Pap. NE-649. Radnor, PA: US. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station. 11 P* 649. doi:10.2737/NE-RP-649.

Weiskittel, Aaron, Christian Kuehne, John Paul McTague, and Mike Oppenheimer. 2016. "Development and Evaluation of an Individual Tree Growth and Yield Model for the Mixed Species Forest of the Adirondacks Region of New York, USA." *Forest Ecosystems* 3 (1). doi:10.1186/s40663-016-0086-3.

Westfall, James A. 2006. "Predicting Past and Future Diameter Growth for Trees in the Northeastern United States." *Canadian Journal of Forest Research 36:1551-1562* 36. https://www.fs.usda.gov/treesearch/pubs/15883.

Westfall, James A., and Kenneth M. Laustsen. 2006. "A Merchantable and Total Height Model for Tree Species in Maine." *Northern Journal of Applied Forestry* 23 (4): 241–49. doi:10.1093/njaf/23.4.241.