

Multilevel Urban Tree Allometric equations

erker

November 7, 2018

Introduction

Multilevel models have been used for decades in tree growth equations (Lappi and Bailey 1988). Indeed the test dataset, "orange", in the statistical programming language, R, is used to demonstrate the fitting of nonlinear multilevel (mixed effects) models ? orange. Multilevel modeling is an attractive approach because it provides a coherent framework to account for the many levels of observation or of groupings in data and to pool information across groups. This paper has two main contributions. First, we demonstrate the use of Stan via the "brms" package in R to fit bayesian nonlinear multilevel models to predict tree diameter growth from age ??. Second, we apply the method to the Urban Tree Database ??. This dataset is the result an over a decade long effort to collect age and size data on thousands of trees in 17 cities across the US. Multilevel modeling has the potential to extract more information from the data and improve predictions compared to the existing modeling approach. Improving predictions of tree size from tree age will improve our ability to predict the important ecosystem services these trees provide urban dwellers.

Stan is a probabilistic programming language for bayesian inference ?. It uses No-U-Turn sampler, an adaptive form of Hamiltonian Monte Carlo sampling, to effectively draw samples from the specified log posterior density. Here, we access Stan via the R package brms ?. brms allows the user to specify the likelihood and priors in syntax similar to the R

package lme4 commonly used for frequentist? multilevel (mixed effects) models. This makes harnessing the power of Stan much simpler and concise because it doesn't require the user to know how to write efficient Stan code and can convert a few lines of R code into many lines of Stan. brms is not as flexible as stan, but still can be used to fit many types of models including nonlinear multilevel regression models, such as ours here. Some of the key advantages of fitting a model in Stan via brms include relatively simple syntax and efficient posterior sampling for multilevel non-linear models. The bayesian approach gives better estimates of parameter uncertainty and provides a formal way to include prior information.

The existing approach to modeling the diameter growth of trees in the urban tree database (UTD) was to fit a separate model for each tree species in each city and test several model forms with different weights and then select the model with the lowest Akaike Information Criterion (AIC). This approach has several limitations, many highlighted in the report. First, while the model form selected provided the lowest AIC, many of the estimates are not biologically realistic (for example they begin to increase sharply at old ages, cubic fits, or decrease at old ages, quadratic fits). Therefore, the researchers discouraged applying the models beyond the range of the data, or sometimes even within the range of data if the estimates were unrealistic. These unrealistic estimates and the inability to extrapolate severely limits managers' ability to predict growth over meaningful time scales (a century rather than a few decades). A second limitations is that some models predict negative diameters, an impossibility. Third, models are only provided for the cities and the species sampled. If a manager wants to predict the diameter growth of a tree species in an unsampled city, the researchers recommend using the model from the reference city in the same climate region. However many of the reference cities are on the border of climate regions and there is known large variability in growth within regions (see figure XX in utd report which is from McPherson... comparing Cheyenne to Ft Collins). Furthermore, if a manager wants to predict growth for an unsampled species or a species that was sampled in a different city it is not obvious which equation/model to use and the additional uncertainty that this introduces

is not quantified.

Our approach addresses the above limitations. First, we use a weibull curve, commonly used in forestry growth equations and biologically realistic, which makes extrapolation to ages outside the data range less fraught. Second, using this sigmoidal curve and modeling diameter with a gamma distribution ensures our estimates of diameter are positive. Third, by modeling the weibull curve parameters as functions of species, city, and climate, we are able to borrow information across cities and across species to provide predictions an associated uncertainty of diameter growth even in even in cities or species with very little or no data.

The sigmoidal curves similar to the weibull have been used before in modeling urban tree diameter growth as a function of age. ?, using an equation of form $y = B_0(1 - \exp(-B_1x))^{B_2}$ predicted DBH from age for healthy trees (12 species, 221 trees total) in Minneapolis and St. Paul, Minnesota. This equation form worked very well (8 out of 12 species had an R^2 over .9), but the trees used were only healthy open grown trees, which is not representative of urban trees generally. Following ?, in an early version of the urban tree database, ? fit the weibull curve to a very small amount of data and adjusted parameters for different locations based on the number of frost free days.

Multilevel modeling in non-urban forestry other work. how is this different? Well the urban part, management practices like topping, pollarding, pruning, drastically alter growth and tree dimensions. Others have discussed the difference between forest and urban/open grown trees a fair bit. The Stan and brms part, these software are new. What about the modeling approach? I don't think many past effects have so many levels, or have things vary by climate. The geographic extent of these urban models is very large. There are a number of papers that have done multilevel modeling for tree growth that I should mention.

A paragraph on the impact of these equations for managing forests to predict/forecast ecosystem services.

Talk more about the bayesian multilevel/hierarchical modeling framework and why it is useful. Show the basis of bayesian hierarchical model.

an early version of the utd equations when there was very little data used the weibull curve ?. They used the number of frost-free days to adjust parameters.

following

? compared the weibull to the logarithm regression and selected logarithm regression for it's higher R^2 (.85)