The forestecology R package for fitting and assessing neighborhood models of the effect of interspecific competition on the growth of trees

Albert Kim¹, David Allen², and Simon Couch³

¹Smith College
²Middlebury College
³Reed College

July 9, 2021

Abstract

1. Neighborhood competition models are powerful tools to measure the effect of interspecific competition. Statistical methods to ease the application of these models are currently lacking. 2. We present the forestecology package providing methods to i) specify neighborhood competition models, ii) evaluate the effect of competitor species identity using permutation tests, and iii) measure model performance using spatial cross-validation. Following Allen (2020), we implement a Bayesian linear regression neighborhood competition model. 3. We demonstrate the package’s functionality using data from the Smithsonian Conservation Biology Institute’s large forest dynamics plot, part of the ForestGEO global network of research sites. Given ForestGEO’s data collection protocols and data formatting standards, the package was designed with cross-site compatibility in mind. We highlight the importance of spatial cross-validation when interpreting model results. 4. The package features i) tidyverse-like structure whereby verb-named functions can be modularly “piped” in sequence, ii) functions with standardized inputs/outputs of simple features ‘sf’ package class, and iii) an S3 object-oriented implementation of the Bayesian linear regression model. These three facts allow for clear articulation of all the steps in the sequence of analysis and easy wrangling and visualization of the geospatial data. Furthermore, while the package only has Bayesian linear regression implemented, the package was designed with extensibility to other methods in mind.

Hosted file

paper.pdf available at https://authorea.com/users/424753/articles/529779-the-forestecology-r-package-for-fitting-and-assessing-neighborhood-models-of-the-effect-of-interspecific-competition-on-the-growth-of-trees
The figure shows a scatter plot with the following labels:

- **fold**: The plot is divided into different sections labeled as 'fold', 'test', 'fold buffer', and 'training'.
- **RMSE (cm y\(^{-1}\))**: The x-axis represents the Root Mean Square Error in centimeters per year, ranging from 0.128 to 0.140.
- **Count**: The y-axis represents the count of data points for each RMSE value.
- **Cross-verified?**: The plot also shows a distribution of data points for cross-verified data, with categories marked as 'FALSE' and 'TRUE'.

The data points are distributed across the plot, indicating the performance of different models or scenarios under various conditions.