RW-2- BIO Model Version 2.0

Estimating aboveground biomass from tree ring measurements

Purpose

Overview

This three-step workflow works to process tree ring measurements and census data into annual estimates of aboveground biomass and diameter at the individual and species levels.

The first step digests and formats the data into appropriate formats for use with the STAN models. It also saves information that will be needed to correctly process the model outputs in the context of the site (e.g. species of individuals, census years, etc.). The two models in the workflow have different STAN files and require slightly different inputs. However, the scripts automatically process the two different types of sites (i.e. census and non-census site) as needed. As a default setting, both models sample only from the year 1900 and on. This decision was made because (1) estimates beyond 1900 will be heavily biased due to the fading record, rendering them unreliable, and (2) the model had convergence issues at select sites when attempting to estimate from the beginning of the ring width dataset. With the current version of the model, starting before 1900 is impractical and unnecessary.

For reference, there are template CSV input files in the repository in the “docs” folder for both the census data file and ring width metadata file. For census data, only diameter values where the tree was recorded as “living” were included. When running the first step, the script will print a warning statement if, based on the data, the predicted initial diameter for any one tree is outside of the range of the model prior distribution for both models. In the case that a warning is printed, one should:

1. Look back at the data and consider, in detail, the individual(s) which flagged the warning. Do the ring width values seem to be entered correctly? Was the measured DBH value for the individual entered in the metadata file correctly?
2. In the case that there are no issues in the data, you can expand the range of the D0 prior distribution within the STAN model, though this will have an impact on model convergence and runtime. Directly adjust the STAN model script as needed in Line 58 (RW + Census Model) or Line 52 (RW only Model).

The second step works to fit and sample from the STAN models. One notable difference between the two models is that the RW only model does not estimate the diameter measurement error, but the RW + Census model does. However, this parameter is necessary for fitting both statistical models. Therefore, if census data are available for a site, we run the RW + Census model first and use the mean estimate of this parameter given by the RW + Census model as the true value for the RW only model. If census data are not available for a site, we use the mean estimate of the parameter from a previously processed site with census data. In the current set-up of the workflow, we are using the estimate from Harvard Forest for sites without census data.

1. The “RW only model” draws 2500 samples that estimate ring width values for all years and individuals where we have RW data. The underlying distribution for each ring width estimate is as follows:

where the former value is particular to the individual and the latter value is particular to the year. These estimated ring width values are then used to determine annual diameter values for each individual by drawing an initial diameter value for each individual and increasing the annual diameter following the estimated increment values. Initial diameter values (D0) are drawn randomly for each individual from the following distribution:

Negative initial diameter values will be removed in the processing step if they occur. All measured ring widths (regardless of coring direction) are considered equally and independently in calculating the likelihood of the drawn sample, which follows the following distribution. We are currently not considering species in the model.

In addition, the measurements of diameter at the time of coring are also used to help calculate the likelihood of the drawn sample following the distribution below. There is only one diameter estimate per individual for this model. Here we see the diameter measurement error parameter that is discussed in the paragraph above in the variance term.

1. The “RW + Census model” draws 2500 samples that estimate ring width values for (1) all years and individuals where we have RW data and (2) individuals and years implied by available census data. For individuals with only census data available, we assume the tree grew by the median measured increment for all individuals and trace back to the first year in which the tree was, at maximum, 10 cm in diameter. We begin to estimate ring width values in this particular year for the individual. By including the census data, we (1) are able to re-introduce biomass lost through mortality, (2) have more diameter data for calculating the likelihood of estimates, (3) can fill in gaps within the nested sampling method, and (4) can estimate diameter measurement error. This model uses the same distributions as the “RW only model.” However, in addition, it uses the census diameter measurements to help calculate the likelihood of the drawn sample following the same Student-t distribution shown above.

The third and final step of the workflow processes the model estimates into annual aboveground biomass and aboveground biomass increment estimates for each individual based on the estimated diameter values from the STAN model. By default, it considers only the last 250 iterations of the model output. If there are census data available for a site, the script automatically completes processing for both the RW only and the RW + Census model. Otherwise, it simply processes just the RW only model.

Upon organizing the diameter estimates from the model into a large array, we use estimated coefficients from Chojnacky et al. (2014) in the following equation to convert diameter (DBH) to aboveground biomass, where S is the species of the individual tree. Chojnacky et al. estimated coefficients B0 and B1 for about 35 different tree taxa groupings. The current CSV file contains coefficients for all PalEON species, as well as most other common species given in Chojnacky et al.

Once biomass has been estimated for all years and individuals, we then iterate through all individuals and remove any data values for a year where the individual’s diameter was estimated to be less than 5 centimeters. The individual will not contribute biomass to the years where it is that small. We also do not include data from the coring year as the ring widths at this time are often incomplete, depending on the time of year when the trees were cored.

For the RW + Census model only, we next perform smoothing for census-indicated mortality. Since the censuses were not conducted annually, we cannot be sure about the specific year in which a tree died if it stops appearing in census records. Therefore, we will stochastically determine the year in which the tree died between the last census in which it was recorded and the first census in which it was not recorded. Biomass and diameter values for the “death year” and all following years are removed from the array. This process is performed for each iteration of the RW + Census model. The model currently assumes any tree with only census data that appeared in the final census record is still alive.

Next, we calculate biomass increment from our biomass array by calculating the difference in biomass for each individual between the current year and the previous year for each relevant model. Then, we restructure the biomass, biomass increment, and diameter arrays into three data frames that contain information on iteration number, species, plot number, model type, year, and individual ID number. These data frames are saved to the site folder. More information is available in the table below.

To create species-level summaries of aboveground biomass, next, we need to first account for the biases introduced by the tree ring sampling design in the RW only model. There are three different methods that are recognized by the workflow: single ring, double-nested, and triple-nested (see figure on next page; PalEON Wiki). In the two nested methods, cores were taken from trees based on their diameter and distance from the plot center. For example, in a double-nested plot, all trees with diameter greater than 10 centimeters were cored within 13 meters from the plot center. Then, from 10 to 20 meters from the plot center, all trees with diameter greater than 20 centimeters were cored. Therefore, since smaller trees are only considered within a fraction of the plot, we can normalize the biomass contribution of smaller trees by dividing each individual’s contribution by the area in which trees of that size were considered. For the single ring design, there is no correction necessary as all class sizes were measured in the plot area. We do not need to perform this step for the RW + Census model since the census helps to fill in gaps in the sampling method for the time period covered by the censuses. We calculate species-level aboveground biomass estimates for each plot and each iteration by summing these normalized biomass contributions of all individuals in each plot for each species. To determine species-level aboveground biomass estimates across the entire site where all plots are of equal size and design, one must only take the average species-level biomass estimate across the plots.



Inputs

The model has been configured such that it requires input of a certain format. For the RW only model, the model requires two types of input. The RW + Census model requires three types of input.

For the RW only model, you will need RWL files that contain the tree ring series for the site in units of 0.001 mm and a CSV file that contains information on the cored individual trees. As common with RWL files, the rownames of the dataset should include the tree ID numbers; any coring direction should be appended to the end of the ID number. The column names should be the year to which the ring width values correspond. All of the RWL files that need to be considered for the model need to be in the site folder “data/raw/rwl.” The following table outlines more specifically what the metadata CSV file should include; a sample CSV file is included in the “doc” folder in the workflow repository. The column names must be exactly the same as shown below.

|  |  |  |
| --- | --- | --- |
| **Data Column** | **Unit** | **Description** |
| ID | string | Identification name for the individual tree which includes the site name, plot number, and tree number; this value must be consistent across all input data files (e.g. “LF1001”) |
| site | string | Plot identification number from which the tree was cored which includes the site name and plot number; this value must be consistent across all input data files (e.g. “LF1”) |
| species | 4-letter USDA code | Recorded species code for individual tree in all capital letters (e.g. QURU); check “data/acronym\_to\_chojnacky\_v0.1.csv” to ensure that the species used in this file will correspond to the correct species in the processing step |
| dbh | cm | Diameter at breast height for individual tree at time of coring |
| distance | m | Distance between individual tree and plot center; this value will be important for correcting for sampling biases. |

For the RW + Census model, you will need the two file types described above. However, you will also need a second CSV file containing additional information on census measurements for trees in the same plots where ring widths were taken. Depending on when the census measurements were taken, there may be trees in this file that do not correspond to trees with ring width data. In the case that there are census data for a tree measured in the ring width dataset, the ID names must be identical between the three data files. **If the ID names do not match up between all three, the model will NOT work correctly.** Here is more specific information on how the census data should be formatted.

|  |  |  |
| --- | --- | --- |
| **Data Column** | **Unit** | **Description** |
| ID | string | Identification name for the individual tree which includes the site name, plot number, and tree number; this value must be consistent across all input data files (e.g. “LF1001”) |
| site | string | Plot identification number where the tree was located which includes the site name and plot number; this value must be consistent across all input data files (e.g. “LF1”) |
| species | 4-letter USDA code | Recorded species code for individual tree in all capital letters (e.g. QURU); check “data/acronym\_to\_chojnacky\_v0.1.csv” to ensure that the species used in this file will correspond to the correct species in the processing step |
| distance | m | Distance between individual tree and plot center; this value will be important for correcting for sampling biases. |
| DXX  (e.g. D69 for 1969 census) | cm | Measured diameter at breast height in the year 19XX or 20XX; the workflow assumes any value for XX that is less than 30 is from the 2000s and all others are from the 1900s. Separate columns are required for each census conducted at the site. |

Outputs

There are two main outputs of the workflow: data files and data figures. The following table outlines the contents of the workflow output data files. The figures and their interpretations are explained in detail below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| File Name | Type | Data Columns | Main Value | Main Unit |
| AGB\_STAN | RDS | iter: iteration number  tree: individual id number  year: year of estimate  model: STAN model used  taxon: species 4-letter code for individual  plot: plot of individual location  type: data type (i.e. ‘ab’)  value: aboveground biomass estimate | individual aboveground biomass | kg |
| AGBI\_STAN | RDS | iter: iteration number  tree: individual id number  year: year of estimate  model: STAN model used  taxon: species 4-letter code for individual  plot: plot of individual location  type: data type (i.e. ‘abi’)  value: aboveground biomass increment estimate | individual aboveground biomass increment | kg/year |
| DBH\_STAN | RDS | iter: iteration number  tree: individual id number  year: year of estimate  model: STAN model used  taxon: species 4-letter code for individual  plot: plot of individual location  type: data type (i.e. ‘dbh’)  value: diameter estimate | individual diameter at breast height | cm |
| AGB\_TAXA\_STAN | RDS | iter: iteration number  year: year of estimate  model: STAN model used  taxon: species 4-letter code for group  plot: plot number  ab: aboveground biomass estimate | species-level aboveground biomass | Mg/ha/plot |

There are 6 types of figures produced by the model workflow, which are all automatically saved to the site’s “figures” folder:

|  |  |  |  |
| --- | --- | --- | --- |
| Figure Type | Figure File Names | Step | Description |
| Histogram of predicted initial diameters for all trees | D0\_histogram\_census.jpg  D0\_histogram.jpg | (1) Build | In addition to the checks within the model script, these figures should be checked to ensure that appropriate initial diameters for each individual will be within the range of the model (marked by the vertical red bars). |
| Line plots demonstrating predicted individual diameter over time | species\_growth\_check.jpg | (1) Build | This figure should be checked to make sure the ring width values were read in correctly prior to running the model (i.e. the tree diameters are experiencing positive growth). |
| Line plots demonstrating estimated individual diameter over time | estimated\_species\_growth\_census.jpg  estimated\_species\_growth.jpg | (2) Run | This figure should be checked to make sure the model estimated ecologically reasonable values for annual diameter growth (i.e. the tree diameters are experiencing positive growth). |
| Estimated total biomass over time | processed\_total\_AGB.jpg | (3) Process | This figure demonstrates the estimated total aboveground biomass over time for each plot measured and for both models (if applicable). |
| Estimated species-level biomass over time | processed\_AGB\_by\_pft.jpg | (3) Process | This figure demonstrates the estimated species-level aboveground biomass over time for each plot measured and for both models (if applicable). |
| 98% biomass percentile | species\_importance.jpg | (3) Process | This figure demonstrates the cumulative biomass contribution of each species to the estimated total site biomass. A horizontal line is included at the 0.98 mark to demonstrate a cut-off for 98% biomass. |

Possible Future Improvements

* Include a figure that shows large discrepancies between predicted diameter based on the ring widths and the census measurements
* Allow the prior distribution range of D0 to be an input parameter to the model
* Apply the sampling correction to data before census data were available

References

Chojnacky, D. C., L. S. Heath, and J. C. Jenkins (2014). Updated generalized biomass equations for North American tree species. Forestry (87): 129-151.

Tree Ring Sampling Designs: <https://paleon.geography.wisc.edu/doku.php/working_groups;npp_protocols>