

APPENDIX II - LandR model details

1. Input data

The LandR Biomass metamodel differs from LANDIS-II in that its workflow consists of self- and on-the-fly parametrization. The only object that the user must supply is a shapefile or raster that defines the study area. If this study area is within Canada, the model is able to parametrize itself, provided that it finds enough data. Our study area in the Northwest Territories was rasterized using a 250 m² grid. Hence, after sourcing the data, most spatial objects were subset and reprojected to match the study area perimeter and resolution.

1.1. Land cover data (Biomass_borealDataPrep module)

Land cover data was used to define the pixels where vegetation and fire dynamics would be simulated (forested pixels), pixels where only fire dynamics were simulated (non-forested, but vegetated pixels - ‘non-forested pixels’ hereafter) and pixels that were excluded entirely from the simulation. We used land cover classes from the 2005 Land Cover of North America (v1) product. Pixels with classes 1 to 15 were included as ‘forested pixels’ and classes 16, 17, 21-25 were included as ‘non-forested pixels’. Recent burns (class 34) and old burns (35) were reclassified into forest pixels, by searching the focal neighbourhood using adjacent forested cover classes (up to a radius of 1250 m from the focal cell). When no forested class was found within this perimeter, the pixel was excluded entirely. Reclassified pixels were omitted from the fitting of statistical models used for parameter estimation, but were assigned predicted values from these models.

1.2. Vegetation data (Biomass_borealDataPrep and Biomass_speciesParameters modules)

Species cover (Biomass_borealDataPrep module)

Species percent cover (% cover) data were obtained from open-source datasets and pre-processed by the Biomass_borealDataPrep module. This module ensures 1) all data use the same geospatial geometries, and 2) that these match the study area. For this work, we used the freely available species % cover rasters derived from MODIS satellite imagery from 2001, prepared by the National Forest Inventory, Natural Resources Canada Canadian Forest Service (NRCan),

(hereafter ‘NFI species data’; Beaudoin et al., 2014). We then selected the species that would be included in the simulation based on a minimum threshold of cover in the study area. The species need to be present in at least 5% of the study area and have at least one pixel with $\geq 10\%$ cover). This resulted in a total of six species or genera: black spruce (*Picea mariana*), white spruce (*Picea glauca*), jack pine (*Pinus banksiana*), paper birch (*Betula papyrifera*), trembling aspen (*Populus tremuloides*), and tamarack (*Larix laricina*).

Stand age and aboveground biomass (Biomass_borealDataPrep module)

Stand age and stand aboveground biomass (hereafter ‘biomass’) were obtained and prepared by the Biomass_borealDataPrep module. Like the NFI species data, both stand age and biomass were rasters derived from MODIS satellite imagery from 2001 prepared by the National Forest Inventory, NRCan (Beaudoin et al., 2014). Biomass_borealDataPrep directly downloads this data and performs a number of data cleaning operations to treat pixels with data inconsistencies. For instance, we detected pixels where species cover was > 0 but biomass was 0, or, inversely, where stand biomass was > 0 but cover was 0. Both age and biomass required fidelity to species % cover, as cover is presumed to be the most accurately estimated variable (especially after the aforementioned overlays). On the other hand, age was considered the least accurate. Hence, where cover equalled 0 and stand age was > 0 , or where age was missing, we imputed species ages using a linear mixed effects model relating age with the interaction between stand biomass (standB), species % cover (cover) and species identity (species), while accounting for the random effect of ecoregion (land cover and ecological zone combination) on intercepts and the slope of the interaction between standB and species.

$$age \sim \log(standB) * cover * species + (\log(standB) * species | ecoregionCode)$$

Predicted ages were subsequently bounded to 0 on the lower limit. Species-specific biomass (B) was estimated for each species present in a given pixel by multiplying its relative cover by standB. Since we used satellite-derived species cover data, the allocation of biomass to deciduous species cohorts was adjusted to reflect the fact that broadleaf canopies will typically have higher cover values for the same amount of biomass as coniferous species. Deciduous species cover was then multiplied by approximately 0.84 before being multiplied by stand biomass to obtain the species

biomass in the pixel. This value comes from a Gaussian generalised linear model (GLM) relating deciduous species biomass (B) with an interaction term between the log of stand age (logAge), standB, *species* and land cover (LC):

$$\log\left(\frac{B}{100}\right) \sim \logAge * \log\left(\frac{standB}{100}\right) * species * LC$$

The model was parameterized using data covering the entire Northwest Territories Province, Canada, and an optimization routine that searched the best conversion factor between deciduous cover and B, by minimizing AIC. The best conversion factor was found by refitting the model on different sets of B values (the response), recalculated by changing deciduous relative cover values between 0.1 and 1.

Permanent sample plot data and simulated species data (Biomass_speciesParameters module)

The Biomass_speciesParameters module used permanent sample plot data (PSP) and simulated data to estimate species traits associated with growth and mortality. The PSP were obtained from the Canadian provinces of British Columbia, Alberta and Saskatchewan, treated for errors and standardized into a single dataset. The data include individual species and diameter at breast height (DBH) measurements for each tree in a plot, as well as stand age. As part of the standardization process, dead trees were removed from the dataset, and a minimum DBH of 10 cm was applied to ensure a consistent measurement cut-off. Tree biomass was then derived from DBH using the model by Lambert et al. (2005) and summed by plot. Lastly, the proportional biomass of each species was calculated for every individual measurement of each plot.

Then, general additive mixed effect models (GAMMs) were used to model the annual biomass of the real species in the PSP dataset. For a given species, observations were excluded from the data if the corresponding species proportional biomass did not exceed 50% (e.g., for *P. tremuloides*, plots were only included if 50% of the biomass was composed of *P. tremuloides*). Measurement year and plot ID were random variables, and the proportional biomass was used to weight observations. Next, a factorial design was used to create over 200,000 combinations of growth and mortality traits (*growthcurve* and *mortalityshape*), longevity, and maximum aboveground net primary productivity (*maxANPP*) of theoretical species. Each theoretical species was then

simulated using LandR Biomass with no reproduction, competition, disturbance, or dispersal, yielding the annual biomass of each theoretical species (i.e. growth curves; see Appendix 2). Lastly, we used log-likelihood to find the most likely combination of species traits by comparing the simulated growth curves and the PSP-derived GAMMs. In LANDIS-II Biomass Succession Extension v3.2, maxB of each species is calculated as 30 times the maxANPP. In LandR Biomass, we calculate maxNPP by multiplying statistically-derived estimates of maxB by 3.33/100 (approx. 30) with 3.33 being the default maxANPP conversion factor, unless adjusted by Biomass_speciesParameters.

Invariant species traits

Most species traits that did not vary spatio-temporally were obtained from available species trait tables used in LANDIS-II:

[https://raw.githubusercontent.com/dcy/LANDIS-](https://raw.githubusercontent.com/dcy/LANDIS-II_IA_generalUseFiles/master/speciesTraits.csv)

[II_IA_generalUseFiles/master/speciesTraits.csv](https://raw.githubusercontent.com/dcy/LANDIS-II_IA_generalUseFiles/master/speciesTraits.csv). Some were then adapted to our study using published literature and statistical models using LandR Biomass_speciesParameters (Barros et al., in prep). Only a few invariant species traits needed to be adjusted “manually” to obtain more realistic successional dynamics in the study area (detailed below).

The LANDIS-II species trait table contains species trait values for each Canadian Ecozone, which we filtered to the Taiga Plains (TP) ecozone. As first runs revealed an excessive recruitment of young cohorts even after several decades of biomass accumulation, we lowered shade tolerance values (shadetolerance) to decrease cohort recruitment as stand biomass increased. We kept relative species shade tolerances equal (*Abies* sp. from 4 to 2.3, *P. glauca* from 3 to 1.6). This aims at reflecting *P. glauca*’s ability to recolonize disturbed sites alongside *Pinus spp.* *Pinus spp* and *Populus spp* already had the lowest value of shade tolerance possible, 1. Mortality (mortalityshape) and growth curve parameters (growthcurve) and adjustment factors for maxANPP (mANPPproportion) and maxB (inflationFactor) were estimated from data where possible using LandR Biomass_speciesParameters, otherwise the default LANDIS-II values were kept.

2. Fire data (FireSense_dataPrep module)

We used fire point data to estimate the average number of fires across the landscape during a reference period. Fire point data were directly downloaded by the FireSense_dataPrep Module

from the National Fire Database available at Canadian Wildland Fire Information System database. The module then filtered the fire occurrences to lightning-caused fires that occurred between 1961-1990. Although this period does not reflect ‘pre-industrial’ conditions (i.e., pre-1940’s), it provides good-confidence data on both fire occurrences and climate, while still being commonly accepted as a baseline with regards to climate change.

3. Data-driven parameter estimation and simulation of regeneration

3.1. Estimation of vegetation parameters

After sourcing and preparing data (see items 1 and 2) the modules `Biomass_borealDataPrep` and `Biomass_speciesParameters` prepare invariant and spatially varying species traits tables by compiling existing trait data and estimating trait values statistically. Below we detail the statistical estimation of some of these traits.

Growth- and mortality curve parameters, maximum biomass and maximum aboveground net primary productivity adjustment factors (`Biomass_speciesParameters` module)

Cohort growth and mortality in LandR Biomass is essentially determined by four parameters: ‘growth curve’, ‘mortality shape’, `maxB` and `maxANPP`. The ‘growth curve’ and ‘mortality shape’ parameters (so called in LANDIS-II Biomass Succession Extension v3.2) modulate the shape of each species’ growth curve and were estimated using `Biomass_speciesParameters`. `maxB` and `maxANPP` determine the maximum biomass and aboveground net primary productivity that a species can have in a particular ecoregion. These were estimated by `Biomass_borealDataPrep` (see below), but adjusted using two factors estimated by `Biomass_speciesParameters`: `maxB` inflation factor (‘inflationFactor’) and `maxANPP` proportion (‘`mANPPproportion`’). `Biomass_speciesParameters` attempts to match the theoretical species curves obtained by varying parameter values (namely, ‘growth curve’, ‘mortality shape’, ‘longevity’ and ‘`mANPPproportion`’) against observed species growth curves from permanent sample plot (PSP) data. Before calculating the observed species growth curves, the PSP data was subset to stand ages below the 95th percent quantile (for each species), as records for larger age classes were limited and constituted statistical outliers. In addition, 50 points were imputed at the origin (age = 0 and biomass = 0) for each species, since very young trees (diameters < 10 cm) were not measured in the original data. This allowed the model to estimate a more realistic intercept value. Growth

curves for each species were then modelled using a generalised additive mixed effects model (GAMM) that related species biomass (B) with stand age ($standAge$), accounting for the random effects of the measurement year ($measureYear$) and plot ($plotID$) on the intercept:

$$B \sim s(standAge) + (\sim 1 | measureYear + plotID)$$

To avoid overfitting, we constrained the smoother on stand age to a maximum smoothing degree of 3 (i.e., 3 knots and a polynomial degree of 2) and a point constraint at 0 that attempted to force the intercept to 0. In addition, B was weighted with respect to species dominance. This consisted in 1) calculating the average biomass of each dominant species, i , (i.e., relative biomass in a plot > 0.5 ; $domSpeciesB_i$), in each plot and measurement year, and 2) dividing the species average biomass by the average biomass across all n species ($mean(domSpeciesB)$):

$$domSpeciesB_i / mean(domSpeciesB_i - n)$$

For the imputed data we used weights equal to 1. Some species did not have enough data to allow for model convergence, thus their growth and mortality curve parameters, together with mANPPproportion and maxB inflation factor, were assigned LANDIS-II default values. After calculating the observed growth curves, Biomass_speciesParameters searched the best matching theoretical growth curves for each species. First, the module selected the theoretical curves with the most similar longevity. The user had the option to constrain the values of the growth and mortality curve parameters, in which case the theoretical curves were also subset to those generated using parameter values within the user's constraints. The maxB inflation factor was then calculated for each of the subset curves, by dividing 5000 (the potential maxB used to simulate theoretical growth curves) by the achieved maximum B in the simulated species growth curves. Finally, the module identified the closest matches between the observed and theoretical species growth curves by maximizing the log-likelihood, after scaling the simulated curves by multiplying the biomass values by the ratio of observed / simulated biomass. Lastly, the module replaces the species trait table with the growth, mortality, and mANPPproportion parameters that were used to simulate the best matching curve. If ties arose, they were caused by different possibilities of mortality values, and handled by selecting the highest mortality parameter value, while the inflation factor was

averaged across the tied curves. Not all species had enough data to build an observed growth curve and find the best matching growth and mortality curve parameters. In this study, observed curves were successfully built for all species.

Maximum biomass and maximum aboveground net primary productivity (Biomass_borealDataPrep and Biomass_speciesParameters modules)

Biomass_borealDataPrep statistically estimates maximum biomass (maxB), maximum aboveground net primary productivity (maxANPP) using observed data. maxB was estimated from a linear mixed effects model reflecting the response of species-specific biomass (B) to the interaction between age (on the log scale, logAge) and species, and the interaction between % cover and species, while accounting for the random effect of ecoregion (the combination of land cover and ecological zone) on the calculated slopes (per species) and intercepts:

$$B \sim \logAge * species + cover * species + (\logAge + cover + species | ecoregion)$$

Coefficients were estimated by maximum likelihood and model fit was calculated as the proportion of explained variance explained by fixed effects only (marginal r^2) and by the entire model (conditional r^2), which was 0.47 and 0.7 respectively. The model was then used to predict maxB per species and ecoregion combination, assuming maximum cover (i.e., 100%) and maximum log-age per species.

Since the statistically estimated maxB reflects field observations, we adjusted the values upwards so that they would relate to potential maximum biomass, rather than observed. To adjust maxB, we multiplied it by the inflation factor estimated for each species using Biomass_speciesParameters (see above). maxANPP was calculated as maxB * mANPPproportion/100, where mANPPproportion defaults to 3.33. This value follows the inversion of the LANDIS-II Biomass Succession Extension v3.2 method to calculate maxB, where maxB = maxANPP * 30 (note that 100/3.33 approx. 30). mANPPproportion was also estimated by Biomass_speciesParameters (see above).

Species establishment probability (Biomass_borealDataPrep module)

Species establishment probability (SEP) was estimated by modelling the probability of observing a given species in each ecoregion. For this, we used a generalized linear mixed model, whereby the probability of occurrence of a species (π) – calculated as the number of pixels with % cover > 0 divided by the total number of pixels, per ecoregion – varied in function of species identity following a binomial distribution (with a logit link), while accounting for the random effect of ecoregion on the intercepts.

$$\text{logit}(\pi) \sim \text{species} + (1 \mid \text{ecoregion})$$

where π is the probability of finding a species in an ecoregion. Model coefficients and fit were calculated in the same way as for the biomass model (see above), resulting in a value of 0.99 for both the marginal r^2 and the conditional r^2 . The fitted values were used as the spatially varying SEP trait values for each species and ecoregion combination.

The modules described here can be found in the following github repositories:

Biomass_borealDataPrep:

https://github.com/PredictiveEcology/Biomass_borealDataPrep/releases/tag/V.2.0_NWT

Biomass_speciesParameters:

https://github.com/PredictiveEcology/Biomass_speciesParameters/releases/tag/V.2.0_NWT

FireSense_dataPrep:

https://github.com/PredictiveEcology/fireSense_dataPrep/releases/tag/V.2.0_NWT

REFERENCES

- Beaudoin, A., Bernier, P. Y., Guindon, L., Villemaire, P., Guo, X. J., Stinson, G., et al. (2014). Mapping attributes of Canada's forests at moderate resolution through kNN and MODIS imagery. *Canadian Journal of Forest Research* 44, 521–532. doi:10.1139/cjfr-2013-0401.