

Edited by: Ceres Barros, Alex M. Chubaty, Ian M. S. Eddy, Eliot J. B. McIntire

LandR Manual

v. 1.0.4



Contents

List of Figures	7
List of Tables	11
Preface	15
1 LandR <i>Biomass_core</i> Module	19
1.1 Module Overview	19
1.1.1 Quick links	19
1.1.2 Summary	20
1.1.3 Links to other modules	21
1.2 Module manual	23
1.2.1 General functioning	23
1.2.2 Initialisation, inputs and parameters	25
1.2.3 List of input objects	32
1.2.4 List of parameters	41
1.2.5 List of outputs	54
1.2.6 Simulation flow and module events	59
1.2.7 Differences between <i>Biomass_core</i> and the LANDIS-II Biomass Succession Extension model (LBSE)	60
1.3 Usage example	69
1.3.1 Set up R libraries	69
1.3.2 Get the module and module dependencies	70
1.3.3 Setup simulation	71

1.3.4	Run simulation	73
1.4	Appendix	75
1.4.1	Tables	75
1.5	References	87
	LandR Data and Calibration Modules	89
2	LandR <i>Biomass_speciesData</i> Module	91
2.1	Module Overview	91
2.1.1	Quick links	91
2.1.2	Module summary	92
2.1.3	Links to other modules	92
2.2	Module manual	92
2.2.1	General functioning	92
2.2.2	List of input objects	94
2.2.3	List of parameters	98
2.2.4	List of outputs	103
2.2.5	Simulation flow and module events	103
2.3	Usage example	105
2.3.1	Load SpaDES and other packages.	105
2.3.2	Set up R libraries	105
2.3.3	Get the module and module dependencies	106
2.3.4	Setup simulation	107
2.3.5	Run module	108
2.4	References	109
3	LandR <i>Biomass_borealDataPrep</i> Module	111
3.1	Module Overview	111
3.1.1	Quick links	111

<i>Contents</i>	5
3.1.2 Summary	112
3.1.3 Links to other modules	112
3.2 Module manual	113
3.2.1 General functioning	113
3.2.2 Data acquisition and treatment	114
3.2.3 Parameter estimation/calibration	119
3.2.4 Aggregating species	126
3.2.5 List of input objects	128
3.2.6 List of parameters	138
3.2.7 List of outputs	153
3.2.8 Simulation flow and module events	156
3.3 Usage example	157
3.4 References	157
4 LandR Biomass_speciesParameters Module	161
4.1 Module Overview	161
4.1.1 Quick links	161
4.1.2 Summary	162
4.1.3 Links to other modules	162
4.2 Module manual	163
4.2.1 General functioning	163
4.2.2 List of input objects	168
4.2.3 List of parameters	175
4.2.4 List of outputs	184
4.2.5 Simulation flow and module events	185
4.3 Usage example	186
4.3.1 Load SpaDES and other packages.	186
4.3.2 Set up R libraries	186

4.3.3	Get the module and module dependencies	187
4.3.4	Setup simulation	187
4.4	References	188
LandR Validation Modules		189
5	LandR Biomass_validationKNN Module	191
5.1	Module Overview	191
5.1.1	Quick links	191
5.1.2	Summary	192
5.1.3	Links to other modules	192
5.2	Module manual	194
5.2.1	General functioning	194
5.2.2	Validation approaches	195
5.2.3	List of input objects	197
5.2.4	List of parameters	209
5.2.5	List of outputs	213
5.2.6	Simulation flow and module events	218
5.3	Usage example	220
5.3.1	Set up R libraries	220
5.3.2	Get the module and module dependencies	220
5.3.3	Setup simulation	221
5.3.4	Run simulation	223
5.3.5	Validate simulation outputs with Biomass_validationKNN	223
5.4	References	227

List of Figures

- 1.1 *Biomass_core* simulates tree cohort growth, mortality, recruitment and dispersal dynamics, as a function of cohort ageing and competition for light (shading) and space, as well as disturbances like fire (simulated using other modules). . 21
- 1.2 Differences in total landscape aboveground biomass when using two different input species orders for the same community. These simulations demonstrate how the sequential calculation of the competition index, combined with a lack of explicit species ordering affect the overall landscape aboveground biomass in time when using different input species orders (see Table reftab:tableLBSEtest1). In order to prevent differences introduced by cohort recruitment, species' ages at sexual maturity were changed to the species' longevity values, and the simulation ran for 75 years to prevent any cohorts from reaching sexual maturity. The bottom panel shows the difference between the two simulations in percentage, calculated as $\frac{Biomass_{order2} - Biomass_{order1}}{Biomass_{order2}} * 100$. 62
- 1.3 Differences in the biomass assigned to new cohorts, summed for each species across pixels, when using two different input species orders for the same community and when the succession time step is 1. These simulations demonstrate how the different summation of total cohort biomass for a succession time step of 1 and the lack of explicit species ordering affect simulation results when changing the species order in the input file (see Table reftab:tableLBSEtest2). Here, initial cohort ages were also set to 1. Values refer to the initial total biomass attributed to each species at the end of year 1. 63

1.4	Hashing design for <i>Biomass_core</i> . In the re-coded <i>Biomass_core</i> , the pixel group map was hashed based on the unique combination of species composition ('community map') and ecolocation map, and associated with a lookup table. The insert in the top-right corner was the original design that linked the map to the lookup table by pixel key.	66
1.5	Visual comparison of simulation outputs for three randomly generated initial communities (left panels) and difference between those outputs (right panels). The % difference between LBSE and <i>Biomass_core</i> were calculated as $\frac{Biomass_{LBSE} - Biomass_{Biomass_core}}{Biomass_{LBSE}} * 100$	68
1.6	Simulation efficiencies of LBSE and <i>Biomass_core</i> with increasing map size, in terms of a) mean running time across repetitions (left y-axis) and the ratio LBSE to <i>Biomass_core</i> running times (right y-axis and blue line), and b) running time scalability as the mean running time per 1000 pixels. .	69
1.7	<i>Biomass_core</i> automatically generates simulation visuals of species dynamics across the landscape in terms of total biomass, number of presences and age and productivity (above), as well as yearly plots of total biomass, productivity, mortality, reproduction and leading species in each pixel (below).	74
2.1	<i>Biomass_speciesData</i> automatically generates a plot of species dominance and number of presences in the study area when '.plotInitialTime=1' is passed as an argument.	109
3.1	Modelling biomass as a linear vs. exponential relationship. a) 'modelBiomass' as 'B ~ logAge * speciesCode + cover * speciesCode + (logAge + cover ecoregionGroup)'. b) 'modelBiomass' as 'logB ~ logAge * speciesCode + cover * speciesCode + (logAge + cover ecoregionGroup)'. Blue dots are marginal mean B values (back-transformed in b) cross ages with confidence intervals as the bars.	122

- 3.2 Thirty years of simulation with 'maxB' values estimated from a 'logB ...' 'biomassModel' (see Fig. reffig:fig-biomassModelLogBtest). The steep increase in such little time is abnormal. 123
- 5.1 *Biomass_validationKNN* automatically generates plots showing a visual comparison between simulated and observed species presences (right) across the landscape, and relative species biomass per pixel (left). 225
- 5.2 A plot of landscape-wide mean absolute deviations (MAD) from (top to bottom) observed mean relative abundance, no. of presences, no. of pixels where the species is dominant and ΔB 226
- 5.3 Diagnostic plot of observed changes in biomass and age ΔB and ΔAge , respectively). 227



List of Tables

1.1	Example of an invariant species traits table (the <code>species</code> table object in the module), with species <i>Abies sp.</i> (<code>Abie_sp</code>), <i>Picea engelmannii</i> (<code>Pice_eng</code>), <i>Picea glauca</i> (<code>Pice_gla</code>), <i>Pinus sp.</i> (<code>Pinu_sp</code>), <i>Populus sp.</i> (<code>Popu_sp</code>) and <i>Pseudotsuga menziesii</i> (<code>Pseu_men</code>). Note that these are theoretical values. (continued below)	28
1.2	Table continues below	28
1.3	Example of a spatio-temporally varying species traits table (the <code>speciesEcoregion</code> table object in the module), with two ecolocations (called <code>ecoregionGroups</code>) and species <i>Abies sp.</i> (<code>Abie_sp</code>), <i>Picea engelmannii</i> (<code>Pice_eng</code>), <i>Picea glauca</i> (<code>Pice_gla</code>), <i>Pinus sp.</i> (<code>Pinu_sp</code>), <i>Populus sp.</i> (<code>Popu_sp</code>) and <i>Pseudotsuga menziesii</i> (<code>Pseu_men</code>). If a simulation runs for 10 year using this table, trait values from year 2 would be used during simulation years 2-10.	30
1.4	Example of a minimum relative biomass table (the <code>minRelativeB</code> table object in the module), with two ecolocations (<code>ecoregionGroups</code>) sharing the same values	31
1.5	Default species probability of germination values used by <i>Biomass_core</i> and <i>Biomass_borealDataPrep</i> . Columns X0-X5 are different site shade levels and each line has the probability of germination for each site shade and species shade tolerance combination.	32
1.6	List of <i>Biomass_core</i> input objects and their description.	36
1.7	List of <i>Biomass_core</i> parameters and their description.	43
1.8	List of <i>Biomass_core</i> output objects and their description.	55

1.9	Input order and processing order (as determined by LBSE) for the same community used to assess the impact of sequential calculation of the competition index, combined with a lack of explicit species ordering. The input order was the order of species in the initial communities table input file. The processing order was the order used in the simulation, which was obtained from <code>Landis-log.txt</code> when <code>CalibrateMode</code> was set to 'yes'. Species starting ages are also shown. (continued below)	76
1.10	Input order and processing order (as determined by LBSE) for the same community used to assess the impact of setting the succession time step to 1, combined with a lack of explicit species ordering. The input order was the order of species in the initial communities table input file. The processing order was the order used in the simulation, which was obtained from <code>Landis-log.txt</code> when <code>CalibrateMode</code> was set to 'yes'. Species starting ages are also shown. (continued below) . .	78
1.11	Randomly generated community combination no. 1 used in the recruitment comparison runs.	80
1.12	Randomly generated community combination no. 2 used in the recruitment comparison runs.	81
1.13	Randomly generated community combination no. 3 used in the recruitment comparison runs.	83
1.14	Invariant species traits table used in comparison runs. (continued below)	85
1.15	Minimum relative biomass table used in comparison runs. XO-5 represent site shade classes from no-shade (0) to maximum shade (5). All ecolocations shared the same values. . .	86
1.16	Probability of germination for species shade tolerance and shade level combinations (called <i>sufficient light</i> table in LBSE and <code>sufficientLight</code> input <code>data.table</code> in LandR <i>Biomass_core</i>) used in comparison runs.	86
1.17	Species ecolocation table used in comparison runs. SEP stands for species establishment probability, <code>maxB</code> for maximum biomass and <code>maxANPP</code> for maximum aboveground net primary productivity. Values were held constant throughout the simulation.	86

2.1	Example of species merging for simulation. Here the user wants to model <i>Abies balsamea</i> , <i>A. lasiocarpa</i> and <i>Pinus contorta</i> as separate species, but all <i>Picea spp.</i> as a genus-level group. For this, all six species are separately identified in the 'KNN' column, so that their % cover layers can be obtained, but in the 'Boreal' column (which defines the naming convention used in the simulation in this example) all <i>Picea spp.</i> have the same name. <i>Biomass_speciesData</i> will merge their % cover data into a single layer by summing their cover per pixel.	94
2.2	List of <i>Biomass_speciesData</i> input objects and their description.	96
2.3	List of <i>Biomass_speciesData</i> parameters and their description.	99
2.4	List of <i>Biomass_speciesData</i> output objects and their description.	103
2.5	List of species cover data downloaded by default by <i>Biomass_speciesData</i>	103
3.1	Example of species merging for simulation. Here the user wants to model <i>Abies balsamea</i> , <i>A. lasiocarpa</i> and <i>Pinus contorta</i> as separate species, but all <i>Picea spp.</i> as a genus-level group. For this, all six species are identified in the 'SCANFI' column, so that their % cover layers can be obtained, but in the 'Boreal' column (which defines the naming convention used in the simulation in this example) all <i>Picea spp.</i> have the same name. <i>Biomass_borealDataPrep</i> will merge their % cover data into a single layer by summing their cover per pixel. (continued below)	127
3.2	List of <i>Biomass_borealDataPrep</i> input objects and their description.	130
3.3	List of <i>Biomass_borealDataPrep</i> parameters and their description.	140
3.4	List of <i>Biomass_borealDataPrep</i> output objects and their description.	154
4.1	List of <i>Biomass_speciesParameters</i> input objects and their description.	170

4.2	List of <i>Biomass_speciesParameters</i> parameters and their description.	177
4.3	List of <i>Biomass_speciesParameters</i> output objects and their description.	184
5.1	List of <i>Biomass_validationKNN</i> input objects and their description.	200
5.2	List of <i>Biomass_validationKNN</i> parameters and their description.	210
5.3	List of <i>Biomass_validationKNN</i> output objects and their description.	214

Preface

Previous released versions of this manual are archived as PDFs and listed at landr-manual.predictiveecology.org/archive/¹.

¹[archive/](#)

LandR



LandR is a collection of SpaDES modules [see 7] aimed at simulating forest dynamics across large spatial scales, while taking into account various disturbances that affect them (e.g., wildfire and climate change), as well as interactions with other components of forest systems, such as bird communities and ungulates (via changes in the habitat of these species), and carbon cycling.

The present manual is a “live” document, in that it grows and changes according to the existing number of LandR modules, as well as their development. In addition, “sub-manuals” may be produced describing particular collections of LandR modules used for a simulation project or a group of similar modules – e.g., a LandR Biomass manual would describe only the LandR modules that are essential for the simulation of the vegetation components of forest succession, excluding disturbances or carbon cycling.

This manual does not cover the SpaDES toolkit, which we use to create and run LandR modules. To learn more about SpaDES go to the SpaDES webpage².

Previous versions:

- LandR Manual v1.0.0³
- LandR Manual v1.0.1⁴
- LandR Manual v1.0.2⁵
- LandR Manual v1.0.3⁶

²<https://spades.predictiveecology.org/>

³<https://github.com/PredictiveEcology/LandR-Manual/raw/main/archive/pdf/LandR-manual-v1.0.0.pdf>

⁴<https://github.com/PredictiveEcology/LandR-Manual/raw/main/archive/pdf/LandR-manual-v1.0.1.pdf>

⁵<https://github.com/PredictiveEcology/LandR-Manual/raw/main/archive/pdf/LandR-manual-v1.0.2.pdf>

⁶<https://github.com/PredictiveEcology/LandR-Manual/raw/main/archive/pdf/LandR-manual-v1.0.3.pdf>



1

LandR Biomass_core Module

Biomass core v. 2.1.0.9002 ¹

Get help Report issues ²

1.0.0.1 Authors:

Yong Luo yluo1@lakeheadu.ca³ [aut], Eliot J B McIntire eliot.mcintire@nrcan-rncan.gc.ca⁴ [aut, cre], Ceres Barros ceres.barros@ubc.ca⁵ [aut], Alex M. Chubaty achubaty@for-cast.ca⁶ [aut], Ian Eddy ian.eddy@nrcan-rncan.gc.ca⁷ [ctb], Jean Marchal jean.d.marchal@gmail.com⁸ [ctb]

This documentation is work in progress. Potential discrepancies and omissions may exist for the time being. If you find any, contact us using the “Get help” link above.

1.1 Module Overview

1.1.1 Quick links

- General functioning
- List of input objects

¹https://github.com/PredictiveEcology/Biomass_core006e37edca05b57a37439b5f8e970ae2516cdc93

²https://github.com/PredictiveEcology/Biomass_core/issues

³<mailto:yluo1@lakeheadu.ca>

⁴<mailto:eliot.mcintire@nrcan-rncan.gc.ca>

⁵<mailto:ceres.barros@ubc.ca>

⁶<mailto:achubaty@for-cast.ca>

⁷<mailto:ian.eddy@nrcan-rncan.gc.ca>

⁸<mailto:jean.d.marchal@gmail.com>

- List of parameters
- List of outputs
- Simulation flow and module events

1.1.2 Summary

LandR *Biomass_core* (hereafter *Biomass_core*) is the core forest succession simulation module of the LandR ecosystem of SpaDES modules [see 7]. It simulates tree cohort ageing, growth, mortality and competition for light resources, as well as seed dispersal (Fig. 1.1), in a spatially explicit manner and using a yearly time step. The model is based on the LANDIS-II Biomass Succession Extension v.3.2.1 [LBSE, 25], with a few changes (see **Differences between *Biomass_core* and LBSE**). Nonetheless, the essential functioning of the succession model still largely follows its LANDIS-II counterpart, and we refer the reader to the corresponding LBSE manual [25] for a detailed reading of the mechanisms implemented in the model.

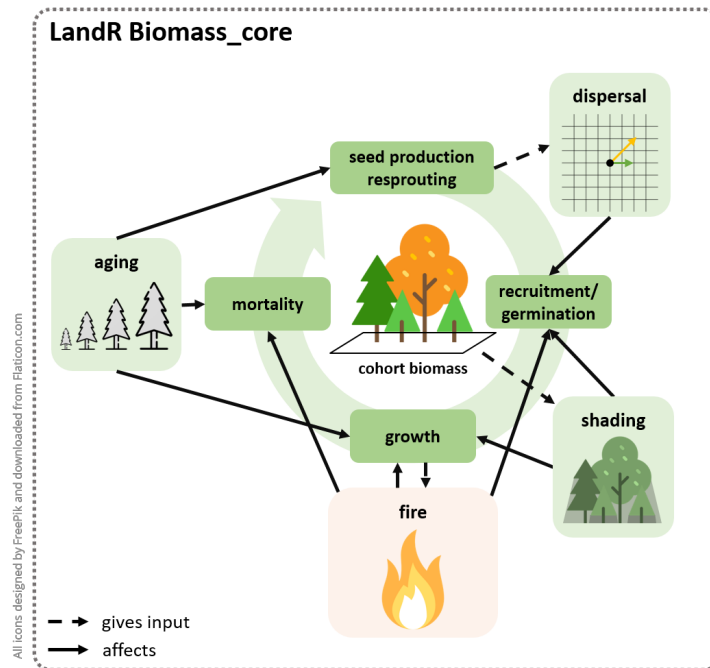


FIGURE 1.1: *Biomass_core* simulates tree cohort growth, mortality, recruitment and dispersal dynamics, as a function of cohort ageing and competition for light (shading) and space, as well as disturbances like fire (simulated using other modules).

1.1.3 Links to other modules

Biomass_core is intended to be used with data/calibration modules, disturbance modules and validation modules, amongst others. The following is a list of the modules most commonly used with *Biomass_core*. For those not yet in the LandR Manual⁹ see the individual module's documentation (.Rmd file) available in its repository.

See here¹⁰ for all available modules and select *Biomass_core* from the drop-down menu to see linkages.

Data and calibration modules:

⁹<https://landr-manual.predictiveecology.org/>

¹⁰https://rpubs.com/PredictiveEcology/LandR_Module_Ecosystem

- *Biomass_speciesData*¹¹: grabs and merges several sources of species cover data, making species percent cover (% cover) layers used by other LandR Biomass modules. Default source data spans the entire Canadian territory;
- *Biomass_borealDataPrep*¹²: prepares all parameters and inputs (including initial landscape conditions) that *Biomass_core* needs to run a realistic simulation. Default values/inputs produced are relevant for boreal forests of Western Canada;
- *Biomass_speciesParameters*¹³: calibrates four-species level traits using permanent sample plot data (i.e., repeated tree biomass measurements) across Western Canada.

Disturbance-related modules:

- *Biomass_regeneration*¹⁴: simulates cohort biomass responses to stand-replacing fires (as in LBSE), including cohort mortality and regeneration through resprouting and/or serotiny;
- *Biomass_regenerationPM*¹⁵: like *Biomass_regeneration*, but allowing partial mortality. Based on the LANDIS-II Dynamic Fuels & Fire System extension [29];
- *fireSense*: climate- and land-cover-sensitive fire model simulating fire ignition, escape and spread processes as a function of climate and land-cover. Includes built-in parameterisation of these processes using climate, land-cover, fire occurrence and fire perimeter data. Requires using *Biomass_regeneration* or *Biomass_regenerationPM*. See modules prefixed “*fireSense_*” at <https://github.com/PredictiveEcology/>;
- *LandMine*¹⁶: wildfire ignition and cover-sensitive wildfire spread model based on a fire return interval input. Requires using *Biomass_regeneration* or *Biomass_regenerationPM*;

¹¹https://github.com/PredictiveEcology/Biomass_speciesData

¹²https://github.com/PredictiveEcology/Biomass_borealDataPrep

¹³https://github.com/PredictiveEcology/Biomass_speciesParameters

¹⁴https://github.com/PredictiveEcology/Biomass_regeneration

¹⁵https://github.com/PredictiveEcology/Biomass_regenerationPM

¹⁶<https://github.com/PredictiveEcology/LandMine>

- *scfm*¹⁷: spatially explicit fire spread module parameterised and modelled as a stochastic three-part process of ignition, escape, and spread. Requires using *Biomass_regeneration* or *Biomass_regenerationPM*.

Validation modules:

- *Biomass_validationKNN*¹⁸: calculates two validation metrics (mean absolute deviation and sum of negative log-likelihoods) on species presences/absences and biomass-related properties across the simulated landscape. By default, it uses an independent dataset of species % cover and stand biomass for 2011, assuming that this is a second snapshot of the landscape.

1.2 Module manual

1.2.1 General functioning

Biomass_core is a forest landscape model based on the LANDIS-II Biomass Succession Extension v.3.2.1 model [LBSE, 25]. It is the core forest succession model of the LandR ecosystem of *SpaDES* modules. Similarly to LBSE, *Biomass_core* simulates changes in tree cohort aboveground biomass (g/m^2) by calculating growth, mortality and recruitment as functions of pixel and species characteristics, competition and disturbances (Fig. 1.1). Note that, by default, cohorts are unique combinations of species and age, but this can be changed via the `cohortDefinitionCols` parameter (see [List of parameters](#)).

Specifically, cohort growth is driven by both invariant (growth shape parameter, `growthcurve`) and spatio-temporally varying species traits (maximum biomass, `maxB`, and maximum annual net primary productivity, `maxANPP`), while background mortality (i.e., not caused by disturbances) depends only on invariant species traits (`longevity` and mortality shape parameter, `mortalityshape`). All these five traits directly influence the realised shape of species growth curves, by determining how fast they grow

¹⁷<https://github.com/PredictiveEcology/scfm>

¹⁸https://github.com/PredictiveEcology/Biomass_validationKNN

(*growthcurve* and *maxANPP*), how soon age mortality starts with respect to longevity (*mortalityshape*) and the biomass a cohort can potentially achieve (*maxB*).

Cohort recruitment is determined by available “space” (i.e., pixel shade), invariant species traits (regeneration mode, *postfireregen*, age at maturity, *sexualmature*, shade tolerance, *shadetolerance*) and a third spatio-temporally varying trait (species establishment probability, *establishprob*, called SEP hereafter). The available “growing space” is calculated as the species’ *maxB* minus the occupied biomass (summed across other cohorts in the pixel). If there is “space”, a cohort can establish from one of three recruitment modes: serotiny, resprouting and germination.

Disturbances (e.g., fire) can cause cohort mortality and trigger post-disturbance regeneration. Two post-disturbance regeneration mechanisms have been implemented, following LBSE: serotiny and resprouting [25]. Post-disturbance mortality and regeneration only occur in response to fire and are simulated in two separate, but interchangeable modules, *Biomass_regeneration* and *Biomass_regenerationPM* that differ with respect to the level of post-fire mortality they simulate (complete or partial mortality, respectively).

Cohort germination (also called cohort establishment) occurs if seeds are available from local sources (the pixel), or via seed dispersal. Seed dispersal can be of three modes: ‘no dispersal’, ‘universal dispersal’ (arguably, only interesting for dummy case studies) or ‘ward dispersal’ [25]. Briefly, the ‘ward dispersal’ algorithm describes a flexible kernel that calculates the probability of a species colonising a neighbour pixel as a function of distance from the source and dispersal-related (and invariant) species traits, and is used by default.

Finally, both germination and regeneration success depend on the species’ probability of germination in a given pixel (**probabilities of germination**).

We refer the reader to Scheller and Miranda [25], Scheller and Domingo [24] and Scheller and Domingo [23] for further details with respect to the above mentioned mechanisms implemented in *Biomass_core*. In a later section of this manual, we highlight existing **differences between *Biomass_core* and LBSE**, together with **comparisons between the two modules**.

1.2.2 Initialisation, inputs and parameters

To initialise and simulate forest dynamics in any given landscape, *Biomass_core* requires a number of inputs and parameters namely:

- **initial cohort biomass and age** values across the landscape;
- **invariant species traits** values;
- **spatio-temporally varying species traits** values (or just spatially-varying);
- **location- (ecolocation-) specific parameters**;
- and the **probabilities of germination** given a species' shade tolerance and site shade.

These are detailed below and in the **full list of input objects**. The *Biomass_borealDataPrep* module manual also provides information about the estimation of many of these traits/inputs from available data, or their adjustment using published values or our best knowledge of boreal forest dynamics in Western Canada.

Unlike the initialisation in LBSE¹⁹, *Biomass_core* initialises the simulation using data-derived initial cohort biomass and age. This information is ideally supplied by data and calibration modules like *Biomass_borealDataPrep* (**Links to other modules**), but *Biomass_core* can also initialise itself using theoretical data.

Similarly, although *Biomass_core* can create all necessary traits and parameters using theoretical values, for realistic simulations these should be provided by data and calibration modules, like *Biomass_borealDataPrep* and *Biomass_speciesParameters*. We advise future users and developers to become familiar with these data modules and then try to create their own modules (or modify existing ones) for their purpose.

1.2.2.1 Initial cohort biomass and age

Initial cohort biomass and age are derived from stand biomass (*biomassMap* raster layer), stand age (*standAgeMap* raster layer) and species % cover

¹⁹in LBSE the initialisation consists in “iterat[ing] the number of time steps equal to the maximum cohort age for each site”, beginning at 0 minus t (t = oldest cohort age) and adding cohorts at the appropriate time until the initial simulation time is reached (0) [25].

(speciesLayers raster layers) data (see Table 1.6) and formatted into the cohortData object. The cohortData table is a central simulation object that tracks the current year's cohort biomass, age, mortality (lost biomass) and aboveground net primary productivity (ANPP) per species and pixel group (pixelGroup). At the start of the simulation, cohortData will not have any values of cohort mortality or ANPP.

Each pixelGroup is a collection of pixels that share the same ecolocation (coded in the ecoregionMap raster layer) and the same cohort composition. By default, an ecolocation is a combination of land-cover and ecological zonation (see ecoregionMap in the full list of inputs) and unique cohort compositions are defined as unique combinations of species, age and biomass. The cohortData table is therefore always associated with the current year's pixelGroupMap raster layer, which provides the spatial location of all pixelGroups, allowing to “spatialise” cohort information and dynamics (e.g., dispersal) on a pixel by pixel basis (see also Hashing).

The user, or another module, may provide initial cohortData and pixelGroupMap objects to start the simulation, or the input objects necessary to produce them: a study area polygon (studyArea), the biomassMap, standAgeMap, speciesLayers and ecoregionMap raster layers (see the list of input objects for more detail).

1.2.2.2 Invariant species traits

These are spatio-temporally constant traits that mostly influence population dynamics (e.g., growth, mortality, dispersal) and responses to fire (fire tolerance and regeneration).

By default, Biomass_core obtains trait values from available LANDIS-II tables (see Table 1.6), but traits can be adjusted/supplied by the user or by other modules. For instance, using Biomass_borealDataPrep will adjust some trait values for Western Canadian boreal forests [e.g., longevity values are adjusted following 5], while using Biomass_speciesParameters calibrates the growthcurve and mortalityshape parameters and estimates two additional species traits (inflationFactor and mANPPproportion) to calibrate maxB and maxANPP (respectively).

Table 1.1 shows an example of a table of invariant species traits. Note that Biomass_core (alone) requires all the columns Table 1.1 in to be present, with the exception of firetolerance, postfireregen, resproutprob, re-

sproutage_min and resproutage_max, which are used by the post-fire regeneration modules (*Biomass_regeneration* and *Biomass_regenerationPM*).

Please see Scheller and Domingo [24, p. 18] and Scheller and Miranda [25, p. 16] for further detail.

TABLE 1.1: Example of an invariant species traits table (the `species` table object in the module), with species *Abies sp.* (`Abie_sp`), *Picea engelmannii* (`Pice_eng`), *Picea glauca* (`Pice_gla`), *Pinus sp.* (`Pinu_sp`), *Populus sp.* (`Popu_sp`) and *Pseudotsuga menziesii* (`Pseu_men`). Note that these are theoretical values. (continued below)

speciesCode	longevity	sexualmature	shadetolerance	firetolerance
Abie_sp	200	20	2.3	1
Pice_eng	460	30	2.1	2
Pice_gla	400	30	1.6	2
Pinu_sp	150	15	1	2
Popu_sp	140	20	1	1
Pseu_men	525	25	2	3

TABLE 1.2: Table continues below

postfireregen	resproutprob	resproutage_min	resproutage_max
none	0	0	0
none	0	0	0
none	0	0	0
serotiny	0	0	0
resprout	0.5	10	70
none	0	0	0

seeddistance_eff	seeddistance_max	mortalityshape	growthcurve
25	100	15	0
30	250	15	1
100	303	15	1
30	100	15	0
200	5000	25	0
100	500	15	1

1.2.2.3 Spatio-temporally varying species traits

These traits vary between species, by ecolocation and, potentially, by year if the year column is not omitted and several years exist (in which case last year's values up to the current simulation year are always used). They are maximum biomass, `maxB`, maximum above-ground net primary productivity, `maxANPP`, and species establishment probability, `SEP` (called `establishprob` in the module). By default, *Biomass_core* assigns theoretical values to these traits, and thus we recommend using *Biomass_borealDataPrep* to obtain realistic trait values derived from data (by default, pertinent for Canadian boreal forest applications), or passing a custom table directly. *Biomass_speciesParameters* further calibrates `maxB` and `maxANPP` by estimating two additional invariant species traits (`inflationFactor` and `maxANPPproportion`; also for Western Canadian forests). See Table 1.3 for an example.

TABLE 1.3: Example of a spatio-temporally varying species traits table (the `speciesEcoregion` table object in the module), with two ecolocations (called `ecoregionGroups`) and species *Abies* sp. (`Abie_sp`), *Picea engelmannii* (`Pice_eng`), *Picea glauca* (`Pice_gla`), *Pinus* sp. (`Pinu_sp`), *Populus* sp. (`Popu_sp`) and *Pseudotsuga menziesii* (`Pseu_men`). If a simulation runs for 10 year using this table, trait values from year 2 would be used during simulation years 2-10.

ecoregionGroup	speciesCode	establishprob	maxB	maxANPP	year
1_03	Abie_sp	1	8567	285	1
1_03	Pice_eng	0.983	10156	305	1
1_03	Popu_sp	0.737	8794	293	1
1_03	Pseu_men	1	17534	132	1
1_09	Abie_sp	0.112	1499	50	1
1_09	Pice_gla	0.302	3143	102	1
1_09	Pinu_sp	0.714	2569	86	1
1_09	Popu_sp	0.607	3292	110	1
1_09	Pseu_men	0.997	6020	45	1
1_03	Abie_sp	0.989	8943	225	2
1_03	Pice_eng	0.985	9000	315	2
1_03	Popu_sp	0.6	8600	273	2
1_03	Pseu_men	1	13534	142	2
1_09	Abie_sp	0.293	2099	45	2
1_09	Pice_gla	0.745	3643	90	2
1_09	Pinu_sp	0.5	2569	80	2

ecoregionGroup	speciesCode	establishprob	maxB	maxANPP	year
1_09	Popu_sp	0.67	3262	111	2
1_09	Pseu_men	1	6300	43	2

1.2.2.4 Ecolocation-specific parameters – minimum relative biomass

Minimum relative biomass (`minRelativeB`) is the only ecolocation-specific parameter used in *Biomass_core*. It is used to determine the shade level in each pixel (i.e., site shade) with respect to the total potential maximum biomass for that pixel (i.e., the sum of all `maxB` values in the pixel's ecolocation). If relative biomass in the stand (with regards to the total potential maximum biomass) is above the minimum relative biomass thresholds, the pixel is assigned that threshold's site shade value [25].

The shade level then influences the germination and regeneration of new cohorts, depending on their shade tolerance (see **Probabilities of germination**).

Site shade varies from X0 (no shade) to X5 (maximum shade). By default, *Biomass_core* uses the same minimum relative biomass threshold values across all ecolocations, adjusted from a publicly available LANDIS-II table²⁰ to better reflect Western Canada boreal forest dynamics (see Table 1.4). *Biomass_borealDataPrep* does the same adjustment by default. As with other inputs, these values can be adjusted by using other modules or by passing user-defined tables.

TABLE 1.4: Example of a minimum relative biomass table (the `minRelativeB` table object in the module), with two ecolocations (`ecoregionGroups`) sharing the same values

ecoregionGroup					
1_03	0.15	0.25	0.5	0.75	0.85
1_09	0.15	0.25	0.5	0.75	0.85

1.2.2.5 Probabilities of germination

A species' probability of germination results from the combination of its shade tolerance level (an invariant species trait in the `species` table; Table 1.1)

²⁰https://github.com/dcyr/LANDIS-II_IA_generalUseFiles

and the site shade – defined by the amount of biomass in the pixel, see **minimum relative biomass parameter** – [25, p. 14]. By default, both *Biomass_core* and *Biomass_borealDataPrep* use a publicly available LANDIS-II table (called `sufficientLight` in the module; Table 1.5).

TABLE 1.5: Default species probability of germination values used by *Biomass_core* and *Biomass_borealDataPrep*. Columns X0-X5 are different site shade levels and each line has the probability of germination for each site shade and species shade tolerance combination.

species shade tolerance	X0	X1	X2	X3	X4	X5
1	1	0	0	0	0	0
2	1	1	0	0	0	0
3	1	1	1	0	0	0
4	1	1	1	1	0	0
5	0	0	1	1	1	1

1.2.2.6 Other module inputs

The remaining module input objects either do not directly influence the basic mechanisms implemented in *Biomass_core* (e.g., `sppColorVect` and `studyAreaReporting` are only used for plotting purposes), are objects that keep track of a property/process in the module (e.g., `lastReg` is a counter of the last year when regeneration occurred), or define the study area for the simulation (e.g., `studyArea` and `rasterToMatch`).

The next section provides a complete list of all input objects, including those already mentioned above.

1.2.3 List of input objects

All of *Biomass_core*'s input objects have (theoretical) defaults that are produced automatically by the module²¹. We suggest that new users run *Biomass_core* by itself supplying only a `studyArea` polygon, before attempting to supply their own or combining *Biomass_core* with other modules. This

²¹usually, default inputs are made when running the `.inputObjects` function (inside the module R script) during the `simInit` call and in the `init` event during the `spades` call – see `?SpaDES.core::events` and `SpaDES.core::simInit`

will enable them to become familiar with all the input objects in a theoretical setting.

Of the inputs listed in Table 1.6, the following are particularly important and deserve special attention:

Spatial layers

- `ecoregionMap` – a raster layer with ecolocation IDs. Note that the term “ecoregion” was inherited from LBSE and kept for consistency with original LBSE code, but we prefer to call them ecolocations to avoid confusion with the ecoregion-level classification of the National Ecological Classification of Canada (NECC)²². Ecolocations group pixels with similar biophysical conditions. By default, we use two levels of grouping in our applications: the first level being an ecological classification such as ecodistricts from the NECC, and the second level is a land-cover classification. Hence, these ecolocations contain relatively coarse scale regional information plus finer scale land cover information. The `ecoregionMap` layer must be defined as a categorical raster, with an associated Raster Attribute Table (RAT; see, e.g., `raster::ratify`). The RAT must contain the columns: ID (the value in the raster layer), `ecoregion` (the first level of grouping) and `ecoregionGroup` (the full ecolocation “name” written as `<firstlevel_secondlevel>`). Note that if creating `ecoregionGroup`’s by combining two raster layers whose values are numeric (as in *Biomass_borealDataPrep*), the group label is a character combination of two numeric grouping levels. For instance, if Natural Ecoregion 2 has land-cover types 1, 2 and 3, the RAT will contain `ID = {1,2,3}`, `ecoregion = {2}` and `ecoregionGroup = {2_1, 2_2, 2_3}`. However, the user is free to use any groupings they wish. Finally, note that all ecolocations (`ecoregionGroup`’s) are should be listed in the `ecoregion` table.
- `rasterToMatch` – a `RasterLayer`, with a given resolution and projection determining the pixels (i.e., non-NA values) where forest dynamics will be simulated. Needs to match `studyArea`. If not supplied, *Biomass_core* attempts to produce it from `studyArea`, using `biomassMap` as the template for spatial resolution and projection.
- `studyArea` – a `SpatialPolygonsDataFrame` with a single polygon determining the where the simulation will take place. This is the only input object that **must be supplied by the user or another module**.

²²<https://open.canada.ca/data/en/dataset/3ef8e8a9-8d05-4fea-a8bf-7f5023d2b6e1>

- `speciesLayers` – layers of species % cover data. Default data currently comes from the Spatialized Canadian National Forest Inventory [SCANFI; Guindon et al. [11], Guindon et al. [12]], a raster-based database of forest characteristics including species covers. The species must match those available in default (or provided) species traits tables (the `species` and `speciesEcoregion` tables).

Species traits and other parameter tables

- `ecoregion` – a `data.table` listing all ecolocation “names” (`ecoregionGroup` column; see `ecoregionMap` above for details) and their state (active – yes – or inactive – no)
- `minRelativeB` – a `data.table` of minimum relative biomass values. See [Ecolocation-specific parameters – minimum relative biomass](#).
- `species` – a `data.table` of [invariant species traits](#).
- `speciesEcoregion` – a `data.table` of [spatio-temporally varying species traits](#).
- `sufficientLight` – a `data.table` defining the probability of germination for a species, given its `shadetolerance` level (see `species` above) and the shade level in the pixel (see `minRelativeB` above). See [Probabilities of germination](#).
- `sppEquiv` – a `data.table` of species name equivalences between various conventions. It must contain the columns *LandR* (species IDs in the LandR format), *EN_generic_short* (short generic species names in English – or any other language – used for plotting), *Type* (type of species, *Conifer* or *Deciduous*, as in “broadleaf”) and *Leading* (same as *EN_generic_short* but with “leading” appended – e.g., “Poplar leading”). See `?LandR::sppEquivalencies_CA` for more information.
- `sppColorVect` – character. A named vector of colours used to plot species dynamics. Should contain one colour per species in the `species` table and, potentially a colour for species mixtures (named “Mixed”). Vector names must follow `species$speciesCode`.
- `sppNameVector` – (OPTIONAL) a character vector of species to be simulated. If provided, *Biomass_core* uses this vector to (attempt to) obtain

speciesLayers for the listed species. If not provided, the user (or another module) can pass a filtered `sppEquiv` table (i.e., containing only the species that are to be simulated). If neither is provided, then *Biomass_core* attempts to use any species for which it finds available species % cover data in the study area.

Cohort-simulation-related objects

- `cohortData` – a `data.table` containing initial cohort information per `pixelGroup` (see `pixelGroupMap` below). This table is updated during the simulation as cohort dynamics are simulated. It must contain the following columns:
 - `pixelGroup` – integer. `pixelGroup` ID. See [Hashing](#).
 - `ecoregionGroup` – character. Ecolocation names. See `ecoregionMap` and `ecoregion` objects above.
 - `speciesCode` – character. Species ID.
 - `age` – integer. Cohort age.
 - `B` – integer. Cohort biomass of the current year in g/m^2 .
 - `mortality` – integer. Cohort dead biomass of the current year in g/m^2 . Usually filled with 0s in initial conditions.
 - `aNPPAct` – integer. Actual aboveground net primary productivity of the current year in g/m^2 . `B` is the result of the previous year's `B` minus the current year's `mortality` plus `aNPPAct`. Usually filled with 0s in initial conditions. See “1.1.3 Cohort growth and ageing” section of Scheller and Miranda [25].
- `pixelGroupMap` – a raster layer with `pixelGroup` IDs per pixel. Pixels are always grouped based on identical `ecoregionGroup`, `speciesCode`, `age` and `B` composition, even if the user supplies other initial groupings (e.g., this is possible in the *Biomass_borealDataPrep* data module).

TABLE 1.6: List of *Biomass_core* input objects and their description.

objectName	objectClass	desc	sourceURL
biomassMap	SpatRaster	Total biomass raster layer in study area (in g/m^2), filtered for pixels covered by cohortData. Only used if <code>P(sim)\$initialBiomassSource == 'biomassMap'</code> , which is currently deactivated.	
cceArgs	list	A list of quoted objects used by the <code>growthAndMortalityDriver</code> <code>calculateClimateEffect</code> function	NA
cohortData	data.table	data.table with cohort-level information on age and biomass, by pixelGroup and ecoregionGroup (i.e., <code>ecoregionGroup</code>). If supplied, it must have the following columns: <code>pixelGroup</code> (integer), <code>ecoregionGroup</code> (factor), <code>speciesCode</code> (factor), <code>B</code> (integer in g/m^2), <code>age</code> (integer in years)	NA
columnsForPixelGroups	character	Optional. If not supplied, will use <code>LandR::columnsForPixelGroups()</code> ; see <code>?LandR::columnsForPixelGroups()</code>	NA

objectName	objectClass	desc	sourceURL
ecoregion	data.table	Ecoregion look up table	https://raw.githubusercontent.com/LANDIS-II-Foundation/Extensions-Succession/master/biomass-succession-archive/trunk/ests/v6.0-2.0/ecoregions.txt
ecoregionMap	SpatRaster	Ecoregion map that has mapcodes matching the ecoregion and speciesEcoregion tables. Defaults to a dummy map matching rasterToMatch with two regions	NA
lastReg	numeric	An internal counter keeping track of when the last regeneration event occurred	NA

objectName	objectClass	desc	sourceURL
pixelGroup Map	SpatRaster	A raster layer with pixelGroup IDs per pixel. Pixels are grouped based on identical ecoregionGroup, speciesCode, age and B composition, even if the user supplies other initial groupings (e.g., via the Biomass_borealDataPrep module.	NA
rasterToMatch	SpatRaster	A raster of the studyArea in the same resolution and projection as biomassMap	NA
species	data.table	A table of invariant species traits with the following trait columns: 'species', 'Area', 'longevity', 'sexualmature', 'shade tolerance', 'firetolerance', 'seeddistance_eff', 'seeddistance_max', 'resprouptprob', 'mortalityshape', 'growthcurve', 'resproutage_min', 'resproutage_max', 'postfireregen', 'wooddecayrate', 'leaflongevity', 'leafLignin', 'hardsoft'. The last seven traits are not used in Biomass_core, and may be omitted. However, this may result in downstream issues with other modules. Default is from Dominic Cyr and Yan Boulanger's project	https://raw.githubusercontent.com/dcyr/LANDIS-II_I_A_generalUsesFiles/master/speciesTraits.csv
speciesEcoregion	data.table	Table of spatially-varying species traits (maxB, maxANPP, establishprob), defined by species and ecoregionGroup (i.e. ecorelocation). Defaults to a dummy table based on dummy data of biomass, age, ecoregion and land cover class	NA

objectName	objectClass	desc	sourceURL
speciesLayers	SpatRaster	cover percentage raster layers by species in Canada species map. Defaults to the Canadian Forestry Service, National Forest Inventory, SCANFI-derived species cover maps from 2020 using a cover threshold of 10 - see https://open.canada.ca/data/en/dataset/18e6a919-53fd-41ce-b4e2-44a9707c52dc for metadata	NA
sppColorVect	character	A named vector of colors to use for plotting. The names must be in <code>sim\$sppEquiv[[sim\$sppEquivCol]]</code> , and should also contain a color for 'Mixed'.	NA
sppEquiv	data.table	Table of species equivalencies. See <code>LandR::sppEquivalencies_CA</code> .	NA
sppNameVector	character	An optional vector of species names to be pulled from <code>sppEquiv</code> . Species names must match <code>P(sim)\$sppEquivCol</code> column in <code>sppEquiv</code> . If not provided, then species will be taken from the entire <code>P(sim)\$sppEquivCol</code> column in <code>sppEquiv</code> . See <code>LandR::sppEquivalencies_CA</code> .	NA
studyArea	SpatVector	Polygon to use as the study area. Must be supplied by the user. Can also be a <code>SpatVector</code> .	NA
studyAreaReporting	SpatVector	multipolygon (typically smaller/unbuffered than <code>studyArea</code>) to use for plotting/reporting. Defaults to <code>studyArea</code> .	NA

objectName	objectClass	desc	sourceURL
sufficient Light	data.frame	Table defining how the species with different shade tolerance respond to stand shade. Default is based on LANDIS-II Biomass Succession v6.2 parameters	https://raw.githubusercontent.com/LANDIS-II-Foundation/Extensions-Succession/master/biomass-succession-archive/trunk/tests/v6.0-2.0/biomass-succession_test.txt
treeFirePixelTableSinceLastDispersal	data.table	3 columns: pixelIndex, pixelGroup, and burnTime. Each row represents a forested pixel that was burned up to and including this year, since last dispersal event, with its corresponding pixelGroup and time it occurred	NA

1.2.4 List of parameters

In addition to the above inputs objects, *Biomass_core* uses several parameters²³ that control aspects like the simulation length, the “succession” time step, plotting and saving intervals, amongst others. Note that a few of these parameters are only relevant when simulating climate effects of cohort growth and mortality, which require also loading the `LandR.CS` R package²⁴ (or another similar package). These are not discussed in detail here, since climate effects are calculated externally to *Biomass_core* in `LandR.CS` functions and thus documented there.

A list of useful parameters and their description is listed below, while the full set of parameters is in Table 1.7. Like with input objects, default values are supplied for all parameters and we suggest the user becomes familiarized with them before attempting any changes. We also note that the “spin-up” and “biomassMap” options for the `initialBiomassSource` parameter are currently deactivated, since *Biomass_core* no longer generates initial cohort biomass conditions using a spin-up based on initial stand age like `LANDIS-II` (“spin-up”), nor does it attempt to fill initial cohort biomasses using `biomassMap`.

Plotting and saving

- `.plots` – activates/deactivates plotting and defines type of plotting (see `?Plots`);
- `.plotInitialTime` – defines when plotting starts;
- `.plotInterval` – defines plotting frequency;
- `.plotMaps` – activates/deactivates map plotting;
- `.saveInitialTime` – defines when saving starts;
- `.saveInterval` – defines saving frequency;

Simulation

- `seedingAlgorithm` – dispersal type (see above);

²³in `SpaDES` lingo parameters are “small” objects, such as an integer or boolean, that can be controlled via the `parameters` argument in `simInit`.

²⁴<https://github.com/ianmseddy/LandR.CS>

- `successionTimestep` – defines frequency of dispersal/local recruitment event (growth and mortality are always yearly);

Other

- `mixedType` – how mixed forest stands are defined;
- `vegLeadingProportion` – relative biomass threshold to consider a species “leading” (i.e., dominant);

TABLE 1.7: List of *Biomass_core* parameters and their description.

paramName	paramClass	default	min	max	paramDesc	cacheKey
calcSummaryBGM	character	end	NA	NA	A character vector describing when to calculate the summary of biomass, growth and mortality. Currently any combination of 5 options is possible: 'start' - as before vegetation succession events, i.e. before dispersal, 'postDisp' - after dispersal, 'postRegen' - after post-disturbance regeneration (currently the same as 'start'), 'postGM' - after growth and mortality, 'postAging' - after aging, 'end' - at the end of vegetation succession events, before plotting and saving. The 'end' option is always active, being also the default option. If NULL, then will skip all summaryBGM related events	TRUE
calibrate	logical	FALSE	NA	NA	Do calibration? Defaults to FALSE	TRUE

paramName	paramClass	default	min	max	paramDesc	cacheKey
cohortDefinitionCols	character	pixelGro....	NA	NA	cohortData columns that determine what constitutes a cohort This parameter should only be modified if additional modules are adding columns to cohortData	TRUE
cutpoint	numeric	1e+10	NA	NA	A numeric scalar indicating how large each chunk of an internal data.table is, when processing by chunks	TRUE
dataYear	numeric	2020	NA	NA	the year for which SCANFI data will be fetched for use with the module. One of 2000, 2010, or 2020, but note that only 2020 is currently supported.	TRUE
initialB	numeric	10	1	NA	Initial biomass values of new age-1 cohorts. If NA or NULL, initial biomass will be calculated as in LANDIS-II Biomass Suc. Extension (see Scheller and Miranda, 2015 or ?LandR::.initiateNewCohorts)	TRUE

44

1 LandR Biomass_core Module

paramName	paramClass	default	min	max	paramDesc	cacheKey
gmcsGrowthLimits	numeric	66.66666....	NA	NA	If using LandR.CS for climate-sensitive growth and mortality, a percentile is used to estimate the effect of climate on growth/mortality (currentClimate/referenceClimate). Upper and lower limits are suggested to circumvent problems caused by very small denominators as well as predictions outside the data range used to generate the model	TRUE
gmcsMinAge	numeric	21	0	NA	If using LandR.CS for climate-sensitive growth and mortality, the minimum age for which to predict climate-sensitive growth and mortality. Young stands (< 30) are poorly represented by the PSP data used to parameterize the model.	TRUE

paramName	paramClass	default	min	max	paramDesc	cacheKey
growthAndMortalityDrivers	character	LandR	NA	NA	Package name where the following functions can be found: calculateClimateEffect, assignClimateEffect (see LandR.CS for climate sensitivity equivalent functions, or leave default if this is not desired)	TRUE
growthInitialTime	numeric	0	NA	NA	Initial time for the growth event to occur	TRUE

paramName	paramClass	default	min	max	paramDesc	cacheKey
initialBiomassSource	character	cohortData	NA	NA	Currently, there are three options: 'spinUp', 'cohortData', 'biomassMap'. If 'spinUp', it will derive biomass by running spinup derived from Landis-II. If 'cohortData', it will be taken from the cohortData object, i.e., it is already correct, by cohort. If 'biomassMap', it will be taken from sim\$biomassMap, divided across species using sim\$speciesLayers percent cover values 'spinUp' uses sim\$standAgeMap as the driver, so biomass is an output . That means it will be unlikely to match any input information about biomass, unless this is set to 'biomassMap', and a sim\$biomassMap is supplied. Only the 'cohortData' option is currently active.	TRUE
keepClimateCols	logical	FALSE	NA	NA	include growth and mortality predictions in cohortData?	TRUE
minCohortBiomass	numeric	9	NA	NA	Cohorts with biomass below this threshold (in g/m^2) are removed. Not a LANDIS-II BSE parameter.	TRUE

paramName	paramClass	default	min	max	paramDesc	cacheKey
mixedType	numeric	2	0	2	How to define mixed stands: 0 for none; 1 for any species admixture; 2 for deciduous > conifer. See ?LandR::vegTypeMapGenerator.	TRUE
plotOverstory	logical	FALSE	NA	NA	swap max age plot with overstory biomass	TRUE
seedingAlgorithm	character	wardDisp....	NA	NA	Choose which seeding algorithm will be used among 'noSeeding' (no horizontal, nor vertical seeding - not in LANDIS-II BSE), 'noDispersal' (no horizontal seeding), 'universalDispersal' (seeds disperse to any pixel), and 'wardDispersal' (default; seeds disperse according to distance and dispersal traits). See Scheller & Miranda (2015) - Biomass Succession extension, v3.2.1 User Guide	TRUE
spinupMortalityfraction	numeric	0.001	NA	NA	Defines the mortality loss fraction in spin up-stage simulation. Only used if P(sim)\$initialBiomassSource == 'biomassMap', which is currently deactivated.	TRUE

48

1 LandR Biomass_core Module

paramName	paramClass	default	min	max	paramDesc	cacheKey
sppEquivCol	character	LandR	NA	NA	The column in <code>sim\$sppEquiv</code> data.table to use as a naming convention during simulation	TRUE
sppEquivPlotCol	character	LandR	NA	NA	The column in <code>sim\$sppEquiv</code> data.table to use as a naming convention for plots	TRUE
successionTimestep	numeric	10	NA	NA	Defines the simulation time step, default is 10 years. Note that growth and mortality always happen on a yearly basis. Cohorts younger than this age will not be included in competitive interactions	TRUE
vegLeadingProportion	numeric	0.75	0	1	A number that defines whether a species is leading for a given pixel. Default: <code>LandR::leadingSpeciesProp()</code> , i.e. option <code>LandR.leadingSpeciesProp</code> , which takes <code>LandR.mixedwoodProp</code> (0.75) unless set. Setting it in one place moves every module and LandR function together.	TRUE
.maxMemory	numeric	5	NA	NA	Maximum amount of memory (in GB) to use for dispersal calculations.	TRUE

paramName	paramClass	default	min	max	paramDesc	cacheKey
.plotInitialTime	numeric	0	NA	NA	Vector of length = 1, describing the simulation time at which the first plot event should occur. To plotting off completely use <code>P(sim)\$plots</code> .	TRUE
.plotInterval	numeric	NA	NA	NA	Defines the plotting time step. If NA, the default, <code>.plotInterval</code> is set to <code>successionTimestep</code> .	TRUE
.plots	character	png	NA	NA	Passed to types in <code>Plots</code> (see <code>?Plots</code>). There are a few plots that are made within this module, if set. Note that plots (or their data) saving will ONLY occur at <code>end(sim)</code> . If NA, plotting is turned off completely (this includes plot saving).	TRUE
.plotMaps	logical	TRUE	NA	NA	Controls whether maps should be plotted or not. Set to FALSE if <code>P(sim)\$plots == NA</code>	TRUE
.plotTransitionField	character	NA	NA	NA	Attribute (field) column in <code>studyAreaReporting</code> to use for defining zones for transition plots. If NA, subpolygons will be aggregated and dissolved for use as a single zone.	TRUE

paramName	paramClass	default	min	max	paramDesc	cacheKey
.plotTransitionNaRm	logical	TRUE	NA	NA	Passed to <code>na.rm</code> in <code>LandR::vegTransitions()</code> for transition plots. If <code>TRUE</code> , pixels with no vegetation type in a transition year are dropped. If <code>FALSE</code> , they are kept and labelled " <code>_NA_</code> "; pixels with no cohorts (e.g., burned and not regenerated), and also pixels in <code>studyAreaReporting</code> that are not simulated (<code>NA</code> in <code>ecoregionMap</code>).	TRUE
.plotTransitionTimes	integer	NA	NA	NA	Simulation times for which transition plots will be built, or <code>NA</code> for none. NOTE: these can be computationally intensive for large landscapes	TRUE
.saveInitialTime	numeric	NA	NA	NA	Vector of length = 1, describing the simulation time at which the first save event should occur. Set to <code>NA</code> if no saving is desired. If not <code>NA</code> , then saving will occur at <code>P(sim)\$saveInitialTime</code> with a frequency equal to <code>P(sim)\$saveInterval</code>	TRUE

paramName	paramClass	default	min	max	paramDesc	cacheKey
.saveInterval	numeric	NA	NA	NA	Defines the saving time step. If NA, the default, .saveInterval is set to $P(sim)\$successionTimestep$.	TRUE
.sslVerify	integer	64	NA	NA	Passed to <code>httr::config(ssl_verifypeer = P(sim)\$sslVerify)</code> when downloading KNN (NFI) datasets. Set to 0L if necessary to bypass checking the SSL certificate (this may be necessary when NFI's website SSL certificate is not correctly configured).	TRUE
.studyAreaName	character	NA	NA	NA	Human-readable name for the study area used. If NA, a hash of <code>studyArea</code> will be used.	TRUE
.useCache	character	.inputOb....	NA	NA	Internal. Can be names of events or the whole module name; these will be cached by SpaDES	TRUE

52

1 LandR Biomass_core Module

paramName	paramClass	default	min	max	paramDesc	cacheKey
.useParallel	ANY	2	NA	NA	Used only in seed dispersal. If numeric, it will be passed to <code>data.table::setDTthreads</code> and should be ≤ 2 ; If <code>TRUE</code> , it will be passed to <code>parallel::makeCluster</code> ; and if a cluster object, it will be passed to <code>parallel::parClusterApplyB</code> .	TRUE
.runName	character	NA	NA	NA	Name for simulation provided by user. Used as a subtitle for plots. <code>NULL</code> is allowed but will result in plots without subtitles.	TRUE

1.2.5 List of outputs

The main outputs of *Biomass_core* are the `cohortData` and `pixelGroupMap` containing cohort information per year (note that they are not saved by default), visual outputs of species level biomass, age and dominance across the landscape and the simulation length, and several maps of stand biomass, mortality and reproductive success (i.e, new biomass) on a yearly basis.

However, any of the objects changed/output by *Biomass_core* (listed in Table 1.8) can be saved via the `outputs` argument in `simInit`²⁵.

²⁵see ?SpaDES.core::outputs

TABLE 1.8: List of *Biomass_core* output objects and their description.

objectName	objectClass	desc
activePixelIndex	integer	Internal use. Keeps track of which pixels are active.
activePixelIndexReporting	integer	Internal use. Keeps track of which pixels are active in the reporting study area.
ANPPMap	SpatRaster	ANPP map at each succession time step (in g/m^2)
biomassMap	SpatRaster	Total biomass raster layer in study area (in g/m^2), filtered for pixels covered by cohortData. Only used if <code>P(sim)\$initialBiomassSource == 'biomassMap'</code> , which is currently deactivated.
cohortData	data.table	<code>data.table</code> with cohort-level information on age, biomass, aboveground primary productivity (year's biomass gain) and mortality (year's biomass loss), by <code>pixelGroup</code> and <code>ecolocation</code> (i.e., <code>ecoregionGroup</code>). Contains at least the following columns: <code>pixelGroup</code> (integer), <code>ecoregionGroup</code> (factor), <code>speciesCode</code> (factor), <code>B</code> (integer in g/m^2), <code>age</code> (integer in years), <code>mortality</code> (integer in g/m^2), <code>anPPAct</code> (integer in g/m^2). May have other columns depending on additional simulated processes (i.e., climate sensitivity; see, e.g., <code>P(sim)\$keepClimateCols</code>).
ecoregion	data.table	Ecoregion look up table

objectName	objectClass	desc
ecoregionMap	SpatRaster	Map with mapcodes match ecoregion table and speciesEcoregion table. Defaults to a dummy map matching rasterToMatch with two regions.
inactivePixelIndex	logical	Internal use. Keeps track of which pixels are inactive.
inactivePixelIndexReporting	integer	Internal use. Keeps track of which pixels are inactive in the reporting study area.
lastFireYear	numeric	Year of the most recent fire.
lastReg	numeric	An internal counter keeping track of when the last regeneration event occurred.
minRelativeB	data.frame	Define the relative biomass cut points to classify stand shade.
mortalityMap	SpatRaster	Map of biomass lost (in g/m^2) at each succession time step.
pixelGroupMap	SpatRaster	Updated community map at each succession time step.
regenerationOutput	data.table	If <code>P(sim)\$calibrate == TRUE</code> , an summary of seed dispersal and germination success (i.e., number of pixels where seeds successfully germinated) per species and year.
reproductionMap	SpatRaster	Regeneration map (biomass gains in g/m^2) at each succession time step
simulatedBiomassMap	SpatRaster	Biomass map at each succession time step (in g/m^2)
simulationOutput	data.table	Contains simulation results by ecoregionGroup (main output)

objectName	objectClass	desc
simulationTreeOutput	data.table	Summary of several characteristics about the stands, derived from <code>cohortData</code>
species	data.table	a table that has species traits such as longevity, shade tolerance, etc. Currently obtained from LANDIS-II Biomass Succession v.6.0-2.0 inputs
speciesEcoregion	data.table	Define the <code>maxANPP</code> , <code>maxB</code> and <code>SEP</code> change with both ecoregion and simulation time.
speciesLayers	SpatRaster	Modified from the input version of this following a call to <code>checkSpeciesTraits()</code>
sppNameVector	character	Modified from the input version of this following a call to <code>sppHarmonize()</code>
spinupOutput	data.table	Spin-up output. Currently deactivated.
sppColorVect	character	A named vector of colors to use for plotting. The names must be in <code>sim\$sppEquiv[[sim\$sppEquivCol]]</code> , and should also contain a color for 'Mixed'.
summaryBySpecies	data.table	The total species biomass (in g/m^2 as in <code>cohortData</code>), average age and aNPP (in g/m^2 as in <code>cohortData</code>), across the landscape (used for plotting and reporting).
summaryBySpecies1	data.table	Number of pixels of each leading vegetation type (used for plotting and reporting).
summaryLandscape	data.table	The averages of total biomass (in tonnes/ha , not g/m^2 like in <code>cohortData</code>), age and aNPP (also in tonnes/ha) across the landscape (used for plotting and reporting).

objectName	objectClass	desc
treedFirePixelTableSinceLastDisp	data.table	3 columns: pixelIndex, pixelGroup, and burnTime. Each row represents a forested pixel that was burned up to and including this year, since last dispersal event, with its corresponding pixelGroup and time it occurred
vegTypeMap	SpatRaster	Map of leading species in each pixel, colored according to sim\$sppColorVect. Species mixtures calculated according to P(sim)\$vegLeadingProportion and P(sim)\$mixedType.

1.2.6 Simulation flow and module events

Biomass_core itself does not simulate disturbances or their effect on vegetation (i.e., post-disturbance mortality and regeneration). Should disturbance and post-disturbance mortality/regeneration modules be used (e.g., *LandMine* and *Biomass_regeneration*), the user should make sure that post-disturbance effects occur *after* the disturbance, but *before* dispersal and background vegetation growth and mortality (simulated in *Biomass_core*). Hence, the disturbance itself should take place either at the very beginning or at the very end of each simulation time step to guarantee that it happens immediately before post-disturbance effects are calculated.

The general flow of *Biomass_core* processes with and without disturbances is:

1. Preparation of necessary objects for the simulation – either by data and calibration modules or by *Biomass_core* itself (during `simInit` and the `init` event²⁶);
2. Disturbances (OPTIONAL) – simulated by a disturbance module (e.g., *LandMine*);
3. Post-disturbance mortality/regeneration (OPTIONAL) – simulated by a regeneration module (e.g., *Biomass_regeneration*);
4. Seed dispersal (every `successionTimestep`; `Dispersal` event):
 - seed dispersal can be a slow process and has been adapted to occur every 10 years (default `successionTimestep`). The user can set it to occur more/less often, with the caveat that if using *Biomass_borealDataPrep* to estimate species establishment probabilities, these values are integrated over 10 years.
 - see Scheller and Domingo [23] for details on dispersal algorithms.
5. Growth and mortality (`mortalityAndGrowth` event):
 - unlike dispersal, growth and mortality always occur time step (year).
 - see Scheller and Mladenoff [27] for further detail.

²⁶`simInit` is a *SpaDES* function that initialises the execution of one or more modules by parsing and checking their code and executing the `.inputObjects` function(s), where the developer provides mechanisms to satisfy each module's expected inputs with default values.

6. Cohort age binning (every `successionTimestep`; `cohortAgeReclassification` event):
 - follows the same frequency as dispersal, collapsing cohorts (i.e., summing their biomass/mortality/aNPP) to ages classes with resolution equal to `successionTimestep`.
 - see Scheller and Miranda [25] for further detail.
7. Summary tables of regeneration (`summaryRegen` event), biomass, age, growth and mortality (`summaryBGM` event);
8. Plots of maps (`plotMaps` event) and averages (`plotAves` and `plotSummaryBySpecies` events);
9. Save outputs (`save` event).

... (repeat 2-9) ...

1.2.7 Differences between *Biomass_core* and the LANDIS-II Biomass Succession Extension model (LBSE)

1.2.7.1 Algorithm changes

Upon porting LBSE into R, we made six minor modifications to the original model's algorithms to better reflect ecological processes. This did not significantly alter the simulation outputs and we note that these changes might also have been implemented in more recent versions of LBSE.

First, for each year and community (i.e., 'pixel group' in *Biomass_core*, see below), LBSE calculates the competition index for a cohort sequentially (i.e., one cohort at a time) after updating the growth and mortality of other cohorts (i.e., their biomass gain and loss, respectively), and with the calculation sequence following cohort age in descending order, but no explicit order of species. This sorting of growth and mortality calculations from oldest to youngest cohorts in LBSE was aimed at capturing size-asymmetric competition between cohorts, under the assumption that older cohorts have priority for growing space given their greater height (Scheller pers. comm.). We felt that within-year sequential growth, death and recruitment may be not ecologically accurate, and that the size-asymmetric competition was being accounted for twice, as the calculation of the competition index already considers the competitive advantage of older cohorts [as shown in the User's Guide,

25]. Hence, in *Biomass_core* growth, mortality, recruitment and the competition index are calculated at the same time across all cohorts and species.

Second, the unknown species-level sorting mechanism contained within LBSE (which changed depending on the species order in the input species list file), led to different simulation results depending on the input species list file (e.g., Table 1.9 and Fig. 1.2). The calculation of competition, growth and mortality for all cohorts at the same time also circumvented this issue.

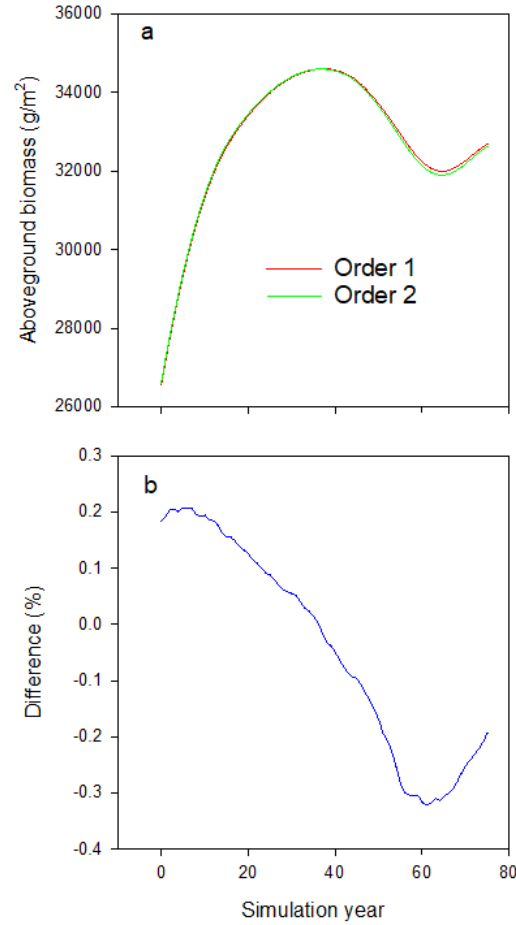


FIGURE 1.2: Differences in total landscape aboveground biomass when using two different input species orders for the same community. These simulations demonstrate how the sequential calculation of the competition index, combined with a lack of explicit species ordering affect the overall landscape aboveground biomass in time when using different input species orders (see Table

reftab:tableLBSEtest1). In order to prevent differences introduced by cohort recruitment, species' ages at sexual maturity were changed to the species' longevity values, and the simulation ran for 75 years to prevent any cohorts from reaching sexual maturity. The bottom panel shows the difference between the two simulations in percentage, calculated as $\frac{Biomass_{order2} - Biomass_{order1}}{Biomass_{order2}} * 100$

Third, in LBSE the calculation of total pixel biomass for the purpose of calculating the initial biomass of a new cohort included the (previously calculated) biomass of other new cohorts when succession time step = 1, but not when time step was > 1. This does not reflect the documentation in the User's Guide, which stated that "*Bsum [total pixel biomass] is the current total biomass for the site (not including other new cohorts)*" [25, p. 4], when the succession time step was set to 1. Additionally, together with the lack of explicit ordering, this generated different results in terms of the biomass assigned to each new cohort (e.g., Table 1.10 and Fig. 1.3). In *Biomass_core* the initial biomass of new cohorts is no longer calculated sequentially (as with competition, growth and mortality), and thus the biomass of new cohorts is never included in the calculation of total pixel biomass.

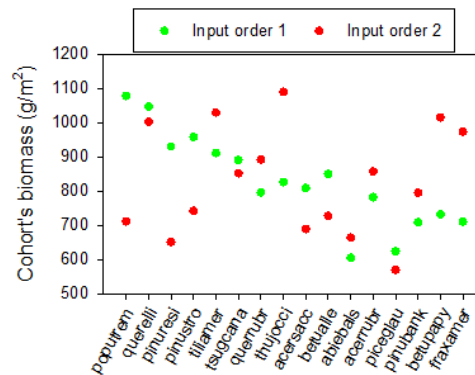


FIGURE 1.3: Differences in the biomass assigned to new cohorts, summed for each species across pixels, when using two different input species orders for the same community and when the succession time step is 1. These simulations demonstrate how the different summation of total cohort biomass for a succession time step of 1 and the lack of explicit species ordering affect simulation results when changing the species order in the input file (see Table reftab:tableLBSEtest2). Here, initial cohort ages were also set to 1. Values refer to the initial total biomass attributed to each species at the end of year 1.

Fourth, in LBSE, serotiny and resprouting could not occur in the same pixel following a fire, with serotiny taking precedence if activated. We understand that this provides an advantage to serotinous species, which could perhaps

be disadvantaged with respect to fast-growing resprouters. However, we feel that it is ecologically more realistic that serotinous and resprouter species be able to both regenerate in a given pixel following a fire and allow the competition between serotinous and resprouting species to arise from species traits. Note that this change was implemented in the *Biomass_regeneration* and *Biomass_regenerationPM* modules, since post-disturbance effects were separated background vegetation dynamics simulated by *Biomass_core*.

Fifth, in *Biomass_core*, species shade tolerance values can have decimal values to allow for finer adjustments of between-species competition.

Sixth, we added a new parameter called `minCohortBiomass`, that allows the user to control cohort removal below a certain threshold of biomass. In some simulation set-ups, we noticed that *Biomass_core* (and LBSE) were able to generate many very small cohorts in the understory that, due to cohort competition, were not able to gain biomass and grow. However, because competition decreases growth but does not increase mortality, these cohorts survived at very low biomass levels until they reached sufficient age to suffer age-related mortality. We felt this is unlikely to be realistic in many cases. By default, this parameter is left at 0 to follow LBSE behaviour (i.e., no cohorts removal based on minimum biomass).

1.2.7.2 Other enhancements

In addition to the sixth changes in growth, mortality and regeneration mentioned above, we enhanced modularity by separating the components that govern vegetation responses to disturbances from *Biomass_core*, and implemented hashing, caching and testing to improve computational efficiency and insure performance.

1.2.7.2.1 Modularity

Unlike in LBSE, post-disturbance effects are not part of *Biomass_core* *per se*, but belong to two separate modules, used interchangeably (*Biomass_regeneration*²⁷ and *Biomass_regenerationPM*²⁸). These need to be loaded and added to the “modules folder” of the project in case the user wants to simulate forest responses to disturbances (only fire disturbances at

²⁷https://github.com/PredictiveEcology/Biomass_regeneration/blob/master/Biomass_regeneration.Rmd

²⁸https://github.com/PredictiveEcology/Biomass_regenerationPM/blob/master/Biomass_regenerationPM.Rmd

the moment). Again, this enables higher flexibility when swapping between different approaches to regeneration.

Climate effects on growth and mortality were also implemented a modular way. The effects of climate on biomass increase (growth) and loss (mortality) were written in functions grouped in two packages. The `LandR` R package contains default, “non-climate-sensitive” functions, while the `LandR.CS` R package contains the functions that simulate climate effects (CS stands for “climate sensitive”). Note that these functions do not simulate actual growth/mortality processes, but estimate modifiers that increase/decrease cohort biomass on top of background growth/mortality. *Biomass_core* uses the `LandR` functions by default (see `growthAndMortalityDrivers` parameter in the [full parameters list](#)). Should the user wish to change how climate effects on growth/mortality are calculated, they can provide new compatible functions (i.e., with the same names, inputs and outputs) via another R package.

1.2.7.2.2 Hashing

Our first strategy to improve simulation efficiency in *Biomass_core* was to use a hashing mechanism [33]. Instead of assigning a key to each pixel in a raster and tracking the simulation for each pixel in a lookup table, we indexed pixels using a *pixelGroup* key that contained unique combinations of ecolocation and community composition (i.e., species, age and biomass composition), and tracked and stored simulation data for each *pixelGroup* (Fig. 1.4). This algorithm was able to ease the computational burden by significantly reducing the size of the lookup table and speeding-up the simulation process. After recruitment and disturbance events, pixels are rehashed into new pixel groups.

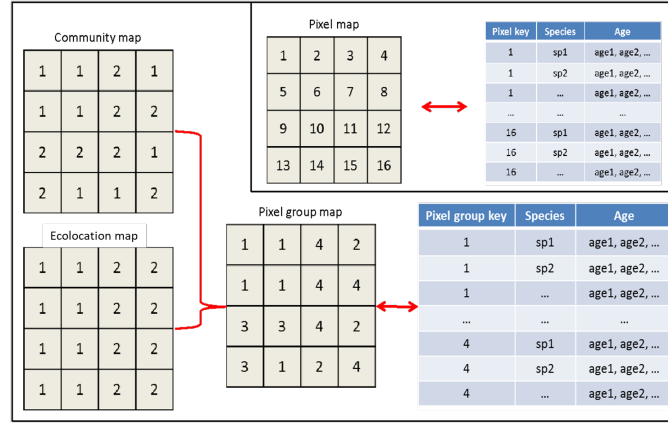


FIGURE 1.4: Hashing design for *Biomass_core*. In the re-coded *Biomass_core*, the pixel group map was hashed based on the unique combination of species composition ('community map') and ecolocation map, and associated with a lookup table. The insert in the top-right corner was the original design that linked the map to the lookup table by pixel key.

1.2.7.2.3 Caching

The second strategy aimed at improving model efficacy was the implementation of caching during data-driven parametrisation and initialisation. Caching automatically archives outputs of a given function to disk (or memory) and reads them back when subsequent calls of this function are given identical inputs. All caching operations were achieved using the reproducible R package [17].

In the current version of *Biomass_core*, the spin-up phase was replaced by data-driven landscape initialisation and many model parameters were derived from data, using data and calibration modules (e.g., *Biomass_borealDataPrep*). To avoid having to repeat data downloads and treatment, statistical estimation of parameters and landscape initialisation every time the simulation is re-run under the same conditions, many of these pre-simulation steps are automatically cached. This means that the pre-simulation phase is significantly faster upon a second call when inputs have not changed (e.g., the input data and parametrisation methods), and when inputs do change only directly affected steps are re-run (see main text for examples). When not using data modules, *Biomass_core* still relies on caching for the preparation of its theoretical inputs.

1.2.7.2.4 Testing

Finally, we implemented code testing to facilitate bug detection by comparing the outputs of functions (etc.) to expected outputs [31]. We built and integrated code tests in *Biomass_core* and across all LandR modules and the LandR R package in the form of assertions, unit tests and integration tests. Assertions and unit tests are run automatically during simulations (but can be turned off) and evaluate individual code components (e.g., one function or an object's class). Integration tests evaluate if several coded processes are integrated correctly and are usually run manually. However, because we embedded assertions within the module code, R package dependencies of *Biomass_core*, such as the LandR R package and *SpaDES*, they also provide a means to test module integration. We also implemented GitHub Actions continuous integration (CI), which routinely test GitHub hosted packages (e.g., LandR) and modules. CRAN-hosted packages (e.g., *SpaDES*) are also automatically tested and checked on CRAN.

Finally, because *Biomass_core* (and all other LandR modules) code is hosted in public GitHub repositories, the module code is subject to the scrutiny of many users, who can identify issues and contribute to improve module code.

1.2.7.3 Performance and accuracy of *Biomass_core* with respect to LBSE

In the recoding of *Biomass_core*, we used integration tests to ensure similar outputs of each demographic process (namely, growth, mortality and recruitment) to the outputs from its counterpart in LBSE. Here, we report the comparisons of the overall simulation (i.e., including all demographic processes) between LBSE and *Biomass_core* using three randomly generated initial communities (Tables 1.11-1.13). The remaining input parameters were taken from a LANDIS-II training course (Tables 1.14-1.17), and contained species attributes information of 16 common tree species in boreal forests and 2 ecoregions. We ran simulations for 1000 years, with a succession time step of 10 and three replicates, which were enough to account for the variability produced by stochastic processes. Seed dispersal was set as “ward dispersal”.

The results suggested that *Biomass_core* had a good agreement with LBSE using the three randomly generated initial communities (Fig. 1.5), with very small deviations for LBSE-generated biomasses. Notably, the mean differences between LBSE and *Biomass_core* were 0.03% (range: -0.01% ~ 0.13%),

0.03% (range: -0.01% ~ 0.11%) and 0.05% (-0.02% ~ 0.15%) for each initial community, respectively (right panels in Fig. 1.5 of this appendix).

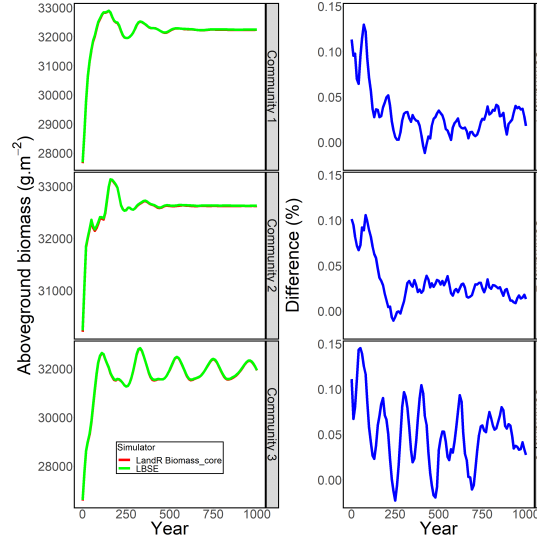


FIGURE 1.5: Visual comparison of simulation outputs for three randomly generated initial communities (left panels) and difference between those outputs (right panels). The % difference between LBSE and *Biomass_core* were calculated as $\frac{Biomass_{LBSE} - Biomass_{Biomass_core}}{Biomass_{LBSE}} * 100$

To examine how running time changed with map size, we ran simulations using maps with increasing number of pixels, from 22,201 to 638,401 pixels. All maps were initialised with a single ecolocation and 7 different communities. Simulations were run for 120 years using a succession time step of 10 and replicated three times. To eliminate the effect of hardware on running time, we used machines that were all purchased at the same time, with equal specifications and running Windows 7. Each simulation ran on 2 CPU threads with a total RAM of 4000 Mb.

For both LBSE and *Biomass_core*, the simulation time increased linearly with number of pixels, but the increase rate was smaller for *Biomass_core* (Fig. 1.6a). This meant that while both models had similar simulation efficiencies in small maps (< 90,000 pixels), as map size increased *Biomass_core* was ~2 times faster than LBSE (maps > 100,000 pixels; Fig. 1.6a). *Biomass_core* also scaled better with map size, as LBSE speeds fluctuated between 19 to 25 sec-

onds per 1,000 pixels across all map sizes, while *Biomass_core* decreased from 21 to 11 seconds per 1,000 pixels from smaller to larger maps (Fig. 1.6b).

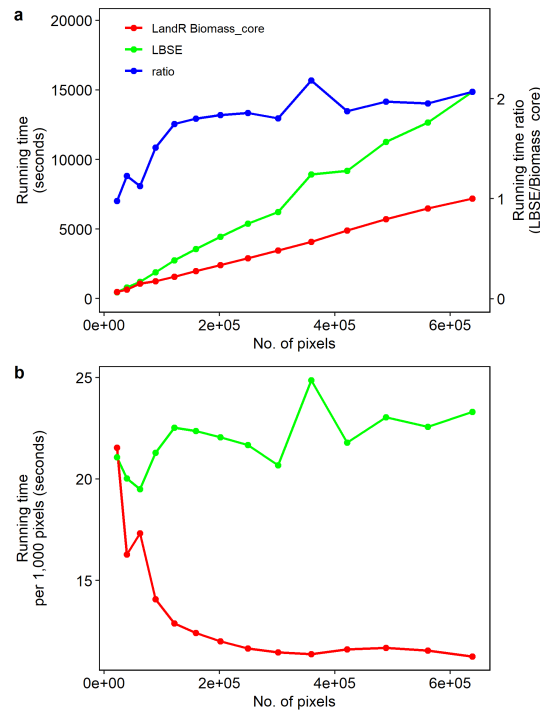


FIGURE 1.6: Simulation efficiencies of LBSE and *Biomass_core* with increasing map size, in terms of a) mean running time across repetitions (left y-axis) and the ratio LBSE to *Biomass_core* running times (right y-axis and blue line), and b) running time scalability as the mean running time per 1000 pixels.

1.3 Usage example

1.3.1 Set up R libraries

```
options(repos = c(CRAN = "https://cloud.r-project.org"))
tempDir <- tempdir()
```

```

pkgPath <- file.path(tempDir, "packages", version$platform,
  paste0(version$major,
    ".", strsplit(version$minor, "[.]")[[1]][1]))
dir.create(pkgPath, recursive = TRUE)
.libPaths(pkgPath, include.site = FALSE)

if (!require(Require, lib.loc = pkgPath)) {
  remotes::install_github(
    "PredictiveEcology/Require@5c44205bf407f613f53546be652a438ef1248147"
  ,
    upgrade = FALSE, force = TRUE)
  library(Require, lib.loc = pkgPath)
}

setLinuxBinaryRepo()

```

1.3.2 Get the module and module dependencies

We can use the `SpaDES.project::getModule` function to download the module to the module folder specified above. Alternatively, see `SpaDES-modules` repository²⁹ to see how to download this and other `SpaDES` modules, or fork/clone from its GitHub repository³⁰ directly.

After downloading the module, it is important to make sure all module R package dependencies are installed in their correct version. `SpaDES.project::packagesInModules` makes a list of necessary packages for all modules in the `paths$modulePath`, and `Require` installs them.

```

Require(
  "PredictiveEcology/SpaDES.project@6d7de6ee12fc967c7c60de44f1aa3b04e6eeb5db"
,
  require = FALSE, upgrade = FALSE, standAlone = TRUE)

```

²⁹<https://github.com/PredictiveEcology/SpaDES-modules>

³⁰https://github.com/PredictiveEcology/Biomass_core/

```

paths <- list(inputPath = normPath(file.path(tempDir,
"inputs")),
  cachePath = normPath(file.path(tempDir, "cache")),
  modulePath = normPath(file.path(tempDir,
    "modules")), outputPath = normPath(file.path(tempDir,
    "outputs")))

SpaDES.project::getModule(modulePath = paths$modulePath,
c("PredictiveEcology/Biomass_core@master"),
  overwrite = TRUE)

## make sure all necessary packages are installed:
outs <- SpaDES.project::packagesInModules(modulePath =
paths$modulePath)
Require(c(unname(unlist(outs)), "SpaDES"), require = FALSE,
standAlone = TRUE)

## load necessary packages
Require(c("SpaDES", "LandR", "reproducible", "pemisc"), upgrade
= FALSE,
  install = FALSE)

```

1.3.3 Setup simulation

Here we setup a simulation in a random study area, using any species within the `LandR::sppEquivalencies_CA` table that can be found there (*Biomass_core* will retrieve species % cover maps and filter present species). We also define the colour coding used for plotting, the type of plots we want to produce and choose to output cohortData tables every year – note that these are not pixel-based, so to “spatialise” results *a posteriori* the `pixelBroupMap` must also be saved.

Please see the lists of **input objects**, **parameters** and **outputs** for more information.

```

times <- list(start = 0, end = 30)

```

```

studyArea <- Cache(randomStudyArea, size = 1e+07) # cache this
so it creates a random one only once on a machine

# Pick the species you want to work with - using the naming
# convention in 'Boreal' column of
# LandR::sppEquivalencies_CA
speciesNameConvention <- "Boreal"
speciesToUse <- c("Pice_Gla", "Popu_Tre", "Pinu_Con")

sppEquiv <- sppEquivalencies_CA[get(speciesNameConvention) %in%
  speciesToUse]
# Assign a colour convention for graphics for each species
sppColorVect <- sppColors(sppEquiv, speciesNameConvention,
  newVals = "Mixed",
  palette = "Set1")

## Usage example
modules <- as.list("Biomass_core")
objects <- list(studyArea = studyArea, sppEquiv = sppEquiv,
  sppColorVect = sppColorVect)

successionTimestep <- 10L

## keep default values for most parameters (omitted from
## this list)
parameters <- list(Biomass_core = list(sppEquivCol =
  successionTimestep = successionTimestep, .plots =
    c("screen",
      "object"), .plotInitialTime = times$start, .plots =
        c("screen",
          "png"), .saveInitialTime = times$start, .useCache =
            "init",
          .useParallel = FALSE))

outputs <- data.frame(expand.grid(objectName = "cohortData",
  saveTime = unique(seq(times$start, times$end, by = 1)),
  eventPriority = 1,
  stringsAsFactors = FALSE))

```

1.3.4 Run simulation

`simInitAndSpades` is a wrapper function that runs both `simInit` (which initialises all modules) and `spades` (which runs all modules, i.e., their events), to which pass all the necessary setup objects created above.

Below, we pass some useful reproducible options that control caching (`"reproducible.useCache"`) and where inputs should be downloaded to (`"reproducible.destinationPath"`).

```
opts <- options(reproducible.useCache = TRUE,
  reproducible.destinationPath = paths$inputPath,
  spades.useRequire = FALSE)
graphics.off()
mySim <- simInitAndSpades(times = times, params = parameters,
  modules = modules, objects = objects, paths = paths, outputs
    = outputs,
  debug = TRUE)
```

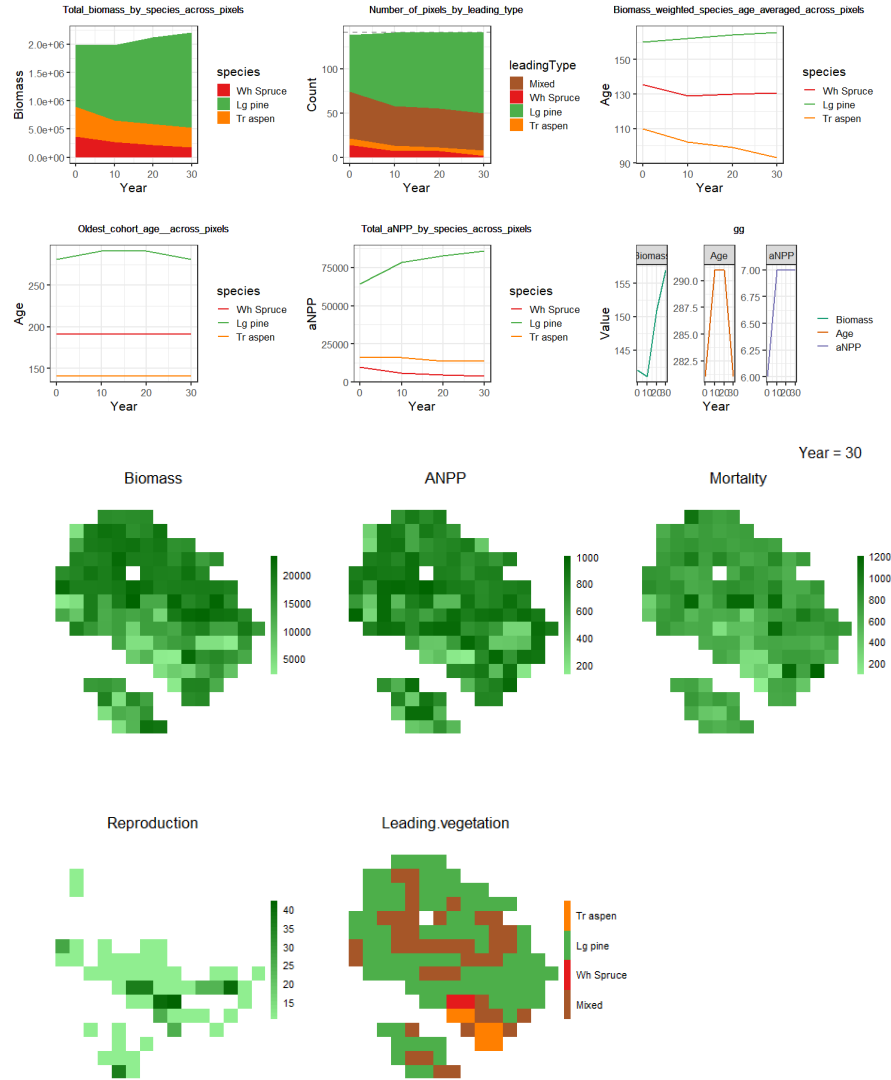


FIGURE 1.7: *Biomass_core* automatically generates simulation visuals of species dynamics across the landscape in terms of total biomass, number of presences and age and productivity (above), as well as yearly plots of total biomass, productivity, mortality, reproduction and leading species in each pixel (below).

1.4 Appendix

1.4.1 Tables

TABLE 1.9: Input order and processing order (as determined by LBSE) for the same community used to assess the impact of sequential calculation of the competition index, combined with a lack of explicit species ordering. The input order was the order of species in the initial communities table input file. The processing order was the order used in the simulation, which was obtained from `Landis-log.txt` when `CalibrateMode` was set to 'yes'. Species starting ages are also shown. (continued below)

Input order 1			Input order 2	
Community	Input order	Age	Processing	Community
1	abiebals	20	poputrem	1
1	acerrubr	20	querelli	1
1	acersacc	20	pinuresi	1
1	betualle	20	pinustro	1
1	betupapy	20	tiliamer	1
1	fraxamer	20	tsugcana	1
1	piceglau	20	querrubr	1
1	pinubank	20	thujocci	1
1	pinuresi	20	acersacc	1
1	pinustro	20	betualle	1
1	poputrem	20	abiebals	1
1	querelli	20	acerrubr	1
1	querrubr	20	piceglau	1
1	thujocci	20	pinubank	1
1	tiliamer	20	betupapy	1
1	tsugcana	20	fraxamer	1

Input order	Age	Processing
pinustro	20	thujocci
poputrem	20	tiliamer
acerrubr	20	querelli
pinubank	20	querrubr
betualle	20	betupapy
piceglau	20	fraxamer
pinuresi	20	tsugcana
acersacc	20	abiebals
querelli	20	acerrubr
querrubr	20	pinubank
thujocci	20	pinustro
tiliamer	20	poputrem
tsugcana	20	pinuresi
abiebals	20	acersacc
betupapy	20	betualle
fraxamer	20	piceglau

TABLE 1.10: Input order and processing order (as determined by LBSE) for the same community used to assess the impact of setting the succession time step to 1, combined with a lack of explicit species ordering. The input order was the order of species in the initial communities table input file. The processing order was the order used in the simulation, which was obtained from `Landis-log.txt` when `CalibrateMode` was set to 'yes'. Species starting ages are also shown. (continued below)

Input order 1			Input order 2	
Community	Input order	Age	Processing	Community
1	abiebals	1	poputrem	1
1	acerrubr	1	querelli	1
1	acersacc	1	pinuresi	1
1	betualle	1	pinustro	1
1	betupapy	1	tiliamer	1
1	fraxamer	1	tsugcana	1
1	piceglau	1	querrubr	1
1	pinubank	1	thujocci	1
1	pinuresi	1	acersacc	1
1	pinustro	1	betualle	1
1	poputrem	1	abiebals	1
1	querelli	1	acerrubr	1
1	querrubr	1	piceglau	1
1	thujocci	1	pinubank	1
1	tiliamer	1	betupapy	1
1	tsugcana	1	fraxamer	1

Input order	Age	Processing
pinustro	1	thujocci
poputrem	1	tiliamer
acerrubr	1	querelli
pinubank	1	querrubr
betualle	1	betupapy
piceglau	1	fraxamer
pinuresi	1	tsugcana
acersacc	1	abiebals
querelli	1	acerrubr
querrubr	1	pinubank
thujocci	1	pinustro
tiliamer	1	poputrem
tsugcana	1	pinuresi
abiebals	1	acersacc
betupapy	1	betualle
fraxamer	1	piceglau

TABLE 1.11: Randomly generated community combination no. 1 used in the recruitment comparison runs.

Community	Species	Age 1	Age 2	Age 3	Age 4	Age 5	Age 6	Age 7
0	betupapy	1	37	45	46	85	NA	NA
0	piceglau	27	73	153	256	270	NA	NA
0	pinustro	157	159	181	220	223	303	307
0	querrubr	80	102	127	152	206	227	NA
1	acerrubr	3	91	126	145	NA	NA	NA
1	acersacc	138	144	276	NA	NA	NA	NA
1	betualle	24	106	136	149	279	NA	NA
1	piceglau	27	67	70	153	NA	NA	NA
1	pinubank	3	10	24	31	71	NA	NA
1	querelli	92	224	234	NA	NA	NA	NA
1	thujocci	73	146	262	NA	NA	NA	NA
2	fraxamer	108	118	137	147	204	NA	NA
2	piceglau	40	128	131	159	174	NA	NA
2	pinustro	78	156	237	245	270	NA	NA
2	querelli	67	97	186	292	NA	NA	NA
2	tiliamer	70	103	121	152	178	180	245
3	acerrubr	5	83	125	126	127	NA	NA
3	pinuresi	1	25	42	49	76	79	103
3	poputrem	4	9	62	NA	NA	NA	NA
3	querelli	101	104	167	226	NA	NA	NA
3	tsugcana	37	135	197	404	405	NA	NA
4	acerrubr	15	29	63	70	105	133	NA
4	piceglau	67	132	189	NA	NA	NA	NA
4	tsugcana	21	26	110	146	341	462	463
5	acerrubr	128	137	145	147	NA	NA	NA
5	acersacc	241	245	261	277	NA	NA	NA
5	querrubr	23	72	120	142	188	NA	NA
5	tiliamer	4	68	98	118	139	197	NA
6	betualle	5	23	31	249	NA	NA	NA
6	pinubank	67	70	89	NA	NA	NA	NA
6	querelli	194	217	257	NA	NA	NA	NA

TABLE 1.12: Randomly generated community combination no. 2 used in the recruitment comparison runs.

Community	Species	Age 1	Age 2	Age 3	Age 4	Age 5	Age 6	Age 7
0	acerrubr	22	26	30	40	47	145	146
0	betualle	23	41	43	120	209	227	270
0	fraxamer	25	90	119	173	185	282	NA
0	pinuresi	48	53	70	121	157	NA	NA
0	pinustro	5	82	126	298	352	NA	NA
0	querrubr	2	30	34	74	77	162	245
1	acerrubr	2	39	43	84	116	127	143
1	pinubank	34	57	75	NA	NA	NA	NA
1	querelli	108	202	218	243	NA	NA	NA
1	querrubr	5	117	131	186	189	246	NA
1	tiliamer	10	19	46	80	133	148	231
1	tsugcana	31	48	190	246	330	NA	NA
2	pinubank	11	37	38	47	67	93	NA
2	querrubr	11	48	57	177	180	228	236
2	tiliamer	28	42	78	79	223	250	NA
2	tsugcana	140	202	372	381	451	NA	NA
3	acersacc	48	107	262	265	NA	NA	NA
3	betupapy	4	12	45	65	83	96	NA
3	poputrem	13	20	37	75	90	NA	NA
3	querelli	72	90	104	115	116	265	278
3	tiliamer	20	21	56	98	237	NA	NA
3	tsugcana	86	224	425	429	NA	NA	NA
4	fraxamer	77	133	181	NA	NA	NA	NA
4	pinustro	13	37	67	220	287	293	375
4	querrubr	27	48	89	97	NA	NA	NA
4	thujocci	91	244	305	390	NA	NA	NA
5	abiebals	86	95	119	121	127	158	NA
5	betualle	83	113	136	161	216	231	NA
5	betupapy	10	38	64	NA	NA	NA	NA
5	piceglau	16	63	70	102	NA	NA	NA
6	acerrubr	8	34	112	NA	NA	NA	NA
6	betupapy	1	31	57	61	74	80	91
6	fraxamer	63	100	108	140	196	294	NA
6	pinubank	15	19	44	47	51	80	NA
6	thujocci	78	146	163	213	214	228	NA

Community	Species	Age 1	Age 2	Age 3	Age 4	Age 5	Age 6	Age 7
6	tsugcana	47	108	387	389	449	NA	NA

TABLE 1.13: Randomly generated community combination no. 3 used in the recruitment comparison runs.

Community	Species	Age 1	Age 2	Age 3	Age 4	Age 5	Age 6	Age 7
0	pinubank	7	26	32	37	48	85	90
0	pinuresi	11	103	109	179	188	197	NA
0	querrubr	89	139	180	206	NA	NA	NA
1	betupapy	36	39	45	49	66	68	NA
1	piceglau	13	165	254	NA	NA	NA	NA
1	pinubank	3	19	54	64	76	NA	NA
1	poputrem	22	59	93	NA	NA	NA	NA
1	thujocci	68	98	274	275	363	378	NA
1	tiliamer	13	20	105	124	248	NA	NA
1	tsugcana	36	90	142	NA	NA	NA	NA
2	fraxamer	11	241	279	NA	NA	NA	NA
2	piceglau	16	42	129	177	200	244	NA
2	pinustro	200	342	384	NA	NA	NA	NA
3	abiebals	31	57	61	92	108	162	183
3	piceglau	126	255	261	267	NA	NA	NA
3	poputrem	28	41	57	NA	NA	NA	NA
3	querrubr	83	91	144	173	184	238	NA
3	thujocci	6	66	68	204	NA	NA	NA
4	fraxamer	12	110	266	270	NA	NA	NA
4	pinustro	174	270	359	379	NA	NA	NA
4	poputrem	4	7	18	24	63	76	NA

Community	Species	Age 1	Age 2	Age 3	Age 4	Age 5	Age 6	Age 7
4	tiliamer	126	136	197	NA	NA	NA	NA
4	tsugcana	49	91	128	194	411	487	NA
5	abiebals	35	53	108	114	147	174	195
5	acerrubr	1	2	101	145	NA	NA	NA
5	pinubank	14	15	38	40	59	69	83
6	acerrubr	4	46	117	NA	NA	NA	NA
6	betualle	36	41	116	213	253	NA	NA
6	betupapy	4	6	76	NA	NA	NA	NA
6	pinuresi	43	68	85	171	NA	NA	NA
6	querrubr	84	86	113	185	193	223	228
6	tiliamer	13	106	181	199	246	NA	NA

84

1 LandR Biomass_core Module

TABLE 1.14: Invariant species traits table used in comparison runs. (continued below)

Species	Longevity	Sexualmature	Shadetolerance	Seeddistance_eff
abiebals	200	25	5	30
acerrubr	150	10	4	100
acersacc	300	40	5	100
betualle	300	40	4	100
betupapy	100	30	2	200
fraxamer	300	30	4	70
piceglau	300	25	3	30
pinubank	100	15	1	20
pinuresi	200	35	2	20
pinustro	400	40	3	60
poputrem	100	20	1	1000
querelli	300	35	2	30
querrubr	250	25	3	30
thujocci	400	30	2	45
tiliamer	250	30	4	30
tsugcana	500	30	5	30

Seeddistance_max	Mortalityshape	Growthcurve
160	10	0.25
200	10	0.25
200	10	0.25
400	10	0.25
5000	10	0.25
140	10	0.25
200	10	0.25
100	10	0.25
275	10	0.25
210	10	0.25
5000	10	0.25
3000	10	0.25
3000	10	0.25
60	10	0.25
120	10	0.25
100	10	0.25

Seeddistance_max	Mortalityshape	Growthcurve
------------------	----------------	-------------

TABLE 1.15: Minimum relative biomass table used in comparison runs. X0-5 represent site shade classes from no-shade (0) to maximum shade (5). All ecolocations shared the same values.

Ecolocation	X0	X1	X2	X3	X4	X5
All	0	0.15	0.25	0.5	0.8	0.95

TABLE 1.16: Probability of germination for species shade tolerance and shade level combinations (called *sufficient light* table in LBSE and `sufficientLight` input data.table in LandR *Biomass_core*) used in comparison runs.

Shadetolerance	0	1	2	3	4	5
1	1	0	0	0	0	0
2	1	1	0	0	0	0
3	1	1	1	0	0	0
4	1	1	1	1	0	0
5	0	0	1	1	1	1

TABLE 1.17: Species ecolocation table used in comparison runs. SEP stands for species establishment probability, `maxB` for maximum biomass and `maxANPP` for maximum aboveground net primary productivity. Values were held constant throughout the simulation.

Ecolocation	Species	SEP	maxANPP	maxB
1	abiebals	0.9	886	26580
1	acerrubr	1	1175	35250
1	acersacc	0.82	1106	33180
1	betualle	0.64	1202	36060
1	betupapy	1	1202	36060
1	fraxamer	0.18	1202	36060
1	piceglau	0.58	969	29070
1	pinubank	1	1130	33900
1	pinuresi	0.56	1017	30510

Ecolocation	Species	SEP	maxANPP	maxB
1	pinustro	0.72	1090	38150
1	poputrem	1	1078	32340
1	querelli	0.96	1096	32880
1	querrubr	0.66	1017	30510
1	thujocci	0.76	1090	32700
1	tiliamer	0.54	1078	32340
1	tsugcana	0.22	1096	32880

1.5 References

- [5] P J Burton and S G Cumming. “Potential effects of climatic change on some western Canadian forests, based on phenological enhancements to a patch model of forest succession”. In: *Water, Air and Soil Pollution* 82 (1995), pp. 401–414.
- [7] Alex M. Chubaty and Eliot J. B. McIntire. *SpaDES: Develop and Run Spatially Explicit Discrete Event Simulation Models*. 2019. URL: <https://CRAN.R-project.org/package=SpaDES>.
- [11] L Guindon et al. *SCANFI: the Spatialized Canadian National Forest Inventory data product*. 2023. DOI: 10.23687/18E6A919-53FD-41CE-B4E2-44A9707C52DC³¹. URL: <https://open.canada.ca/data/en/dataset/18e6a919-53fd-41ce-b4e2-44a9707c52dc> (visited on 09/26/2025).
- [12] Luc Guindon et al. “A new approach for spatializing the Canadian National Forest Inventory (SCANFI) using Landsat dense time series”. en. In: *Canadian Journal of Forest Research* 54.7 (July 2024), pp. 793–815. ISSN: 0045-5067, 1208-6037. DOI: 10.1139/cjfr-2023-0118³². URL: <https://cdnsiencepub.com/doi/10.1139/cjfr-2023-0118> (visited on 09/26/2025).
- [17] Eliot J. B. McIntire and Alex M. Chubaty. *reproducible: A Set of Tools that Enhance Reproducibility Beyond Package Management*. 2020. URL: <https://reproducible.predictiveecology.org,%20https://github.com/PredictiveEcology/reproducible>.

³¹<https://doi.org/10.23687/18E6A919-53FD-41CE-B4E2-44A9707C52DC>

³²<https://doi.org/10.1139/cjfr-2023-0118>

- [23] Robert M. Scheller and James B. Domingo. *LANDIS-II Model v6.0 – Conceptual Description*. Apr. 20, 2012.
- [24] Robert M. Scheller and James B. Domingo. *LANDIS-II Model v6.0 – User Guide*. July 19, 2011.
- [25] Robert M. Scheller and Brian R. Miranda. *LANDIS-II Biomass Succession v3.2 Extension – User Guide*. 2015.
- [27] Robert M. Scheller and David J. Mladenoff. “A forest growth and biomass module for a landscape simulation model, LANDIS: design, validation, and application”. In: *Ecological Modelling* 180.1 (Dec. 2004), pp. 211–229. DOI: 10.1016/j.ecolmodel.2004.01.022³³. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0304380004003837>.
- [29] Brian R. Sturtevant et al. *LANDIS-II Dynamic Fire System Extension v3.0 – User Guide*. 2018.
- [31] Hadley Wickham. “testthat: Get Started with Testing”. In: *The R Journal* 3.1 (2011), p. 5. DOI: 10.32614/RJ-2011-002³⁴. URL: <https://journal.r-project.org/archive/2011/RJ-2011-002/index.html>.
- [33] Jian Yang et al. “An innovative computer design for modeling forest landscape change in very large spatial extents with fine resolutions”. In: *Ecological Modelling* 222.15 (Aug. 10, 2011), pp. 2623–2630. DOI: 10.1016/j.ecolmodel.2011.04.032³⁵. URL: <https://www.sciencedirect.com/science/article/pii/S0304380011002651>.

³³<https://doi.org/10.1016/j.ecolmodel.2004.01.022>

³⁴<https://doi.org/10.32614/RJ-2011-002>

³⁵<https://doi.org/10.1016/j.ecolmodel.2011.04.032>

LandR Data and Calibration Modules

The LandR ecosystem of SpaDES modules has a variety of data and/or calibration modules that are used to obtain and pre-process input data, as well as estimate input parameters required by the core forest landscape simulation module *Biomass_core*. These modules are presented in the next chapters.



2

LandR Biomass_speciesData Module

Biomass speciesData v. 1.1.0.9000 ¹

Get help Report issues ²

2.0.0.1 Authors:

Ceres Barros ceres.barros@ubc.ca³ [aut, cre], Eliot J B McIntire eliot.mcintire@nrcan-rncan.gc.ca⁴ [aut], Alex M. Chubaty achubaty@for-cast.ca⁵ [aut]

This documentation is work in progress. Potential discrepancies and omissions may exist for the time being. If you find any, contact us using the “Get help” link above.

2.1 Module Overview

2.1.1 Quick links

- General functioning
- List of input objects
- List of parameters
- List of outputs

¹https://github.com/PredictiveEcology/Biomass_speciesData68c88fc34089fb20786dafc9cc4a6fd5f7387f49

²https://github.com/PredictiveEcology/Biomass_speciesData/issues

³<mailto:ceres.barros@ubc.ca>

⁴<mailto:eliot.mcintire@nrcan-rncan.gc.ca>

⁵<mailto:achubaty@for-cast.ca>

- **Simulation flow and module events**

2.1.2 Module summary

LandR *Biomass_speciesData* (hereafter *Biomass_speciesData*) downloads and pre-processes species percent (% cover) data layers used by other LandR data modules (e.g., *Biomass_borealDataPrep*) and by the LandR forest simulation module *Biomass_core*.

2.1.3 Links to other modules

Biomass_speciesData is intended to be used with any LandR modules that require species % cover raster layers (see examples below). See here⁶ for all available modules in the LandR ecosystem and select *Biomass_speciesData* from the drop-down menu to see potential linkages.

- *Biomass_borealDataPrep*⁷: prepares all parameters and inputs (including initial landscape conditions) that *Biomass_core* needs to run a realistic simulation. Default values/inputs produced are relevant for boreal forests of Western Canada. Used downstream from *Biomass_speciesData*;
- *Biomass_core*⁸: core forest dynamics simulation module. Used downstream from *Biomass_speciesData*.

2.2 Module manual

2.2.1 General functioning

Biomass_speciesData accesses and processes species % cover data for the parametrisation and initialisation of LandR *Biomass_core*. This module ensures 1) that all data use the same geospatial geometries and 2) that these are correctly re-projected to the study area used for parametrisation

⁶https://rpubs.com/PredictiveEcology/LandR_Module_Ecosystem

⁷https://github.com/PredictiveEcology/Biomass_borealDataPrep

⁸https://github.com/PredictiveEcology/Biomass_core

(studyAreaLarge polygon), and 3) attempts to sequentially fill-in and replace the lowest quality data with higher quality data when several data sources are used. Its primary output is a RasterStack of species % cover, with each layer corresponding to a species.

Currently, the module can access the following datasets: - Canadian National Forest Inventory (NFI) forest attributes kNN dataset [kNN; Beaudoin et al. [3], Beaudoin et al. [4]]; - Common Attribute Schema for Forest Resource Inventories [CAFRI (v4); Cosco [9]], - Spatialized Canadian National Forest Inventory [SCANFI; Guindon et al. [11], Guindon et al. [12]]; - National Terrestrial Ecosystem Monitoring System for Canada Annual Tree Species [NTEMS; Hermosilla et al. [13]]; - Manitoba Forest Resource Inventory (MBFRI); - Ontario Forest Resource Inventory (ONFRI); - historical dataset specific to Western Canada compiled by Paul Pickell and Nicholas Coops [22] as part of the LandWeb project; - “current conditions” dataset specific to Western Canada compiled from forest resource inventories as part of the LandWeb project (ForestInventory);

However, **only the NFI kNN data are freely available** and access to the other datasets must be granted by module developers and data owners, and requires a Google account. Nevertheless, the module is flexible enough that any user can use it to process additional datasets, provided that an adequate R function is passed to the module (see types parameter details in the [list of parameters](#)).

When multiple data sources are used, the module will replace lower quality data with higher quality data following the order specified in the types parameter.

When multiple species of a given data source are to be grouped, % cover is summed across species of the same group within each pixel. Please see the sppEquiv in the [list of input objects](#) for information on how to define species groups.

The module can also exclude species % cover layers if they don't have a minimum % cover value in at least one pixel. The user should still inspect where species is deemed present (e.g., in how many pixels in total), as it is possible that some datasets only have a few pixels where the species is present, but with reported high % cover. In this case, the user may choose to exclude these species *a posteriori*. The summary plot automatically shown by *Biomass_speciesData* can help diagnose whether certain species are present in very few pixels (see Fig. 2.1).

2.2.2 List of input objects

Below is the full list of input objects that *Biomass_speciesData* requires (Table 2.2). Of these, the only input that **must** be provided (i.e., *Biomass_speciesData* does not have a default for) is *studyAreaLarge*.

Of the inputs in Table 2.2, the following are particularly important and deserve special attention:

- *studyAreaLarge* – the polygon defining the area for which species cover data are desired. It can be larger (but never smaller) than the study area used in the simulation of forest dynamics (i.e., *studyArea* object in *Biomass_core*), in which case it should fully cover it.
- *sppEquiv* – a table of correspondences between different species naming conventions. This table is used across several LandR modules, including *Biomass_core*. It is particularly important here because it will determine whether and which species (and their cover layers) are merged. For instance, if the user wishes to simulate a generic *Picea spp.* that includes, *Picea glauca*, *Picea mariana* and *Picea engelmannii*, they will need to provide these three species names in the data column (e.g., KNN if obtaining forest attribute kNN data layers from the National Forest Inventory), but the same name (e.g., “Pice_Spp”) in the column chosen for the naming convention used throughout the simulation (defined by the *sppEquivCol* parameter). See Table 2.1 for an example.

TABLE 2.1: Example of species merging for simulation. Here the user wants to model *Abies balsamea*, *A. lasiocarpa* and *Pinus contorta* as separate species, but all *Picea spp.* as a genus-level group. For this, all six species are separately identified in the ‘KNN’ column, so that their % cover layers can be obtained, but in the ‘Boreal’ column (which defines the naming convention used in the simulation in this example) all *Picea spp.* have the same name. *Biomass_speciesData* will merge their % cover data into a single layer by summing their cover per pixel.

Species	KNN	Boreal	Modelled as
<i>Abies balsamea</i>	Abie_Bal	Abie_Bal	<i>Abies balsamea</i>
<i>Abies lasiocarpa</i>	Abie_Las	Abie_Las	<i>Abies lasiocarpa</i>
<i>Picea engelmannii</i>	Pice_Eng	Pice_Spp	<i>Picea spp.</i>

Species	KNN	Boreal	Modelled as
<i>Picea engelmannii</i> x <i>glauca</i>	Pice_Eng_Gla	Pice_Spp	<i>Picea</i> spp.
<i>Picea glauca</i>	Pice_Gla	Pice_Spp	<i>Picea</i> spp.
<i>Picea mariana</i>	Pice_Mar	Pice_Spp	<i>Picea</i> spp.
<i>Pinus contorta</i> var. <i>latifolia</i>	Pinu_Con	Pinu_Con	<i>Pinus contorta</i> var. <i>latifolia</i>

TABLE 2.2: List of *Biomass_speciesData* input objects and their description.

objectName	objectClass	desc	sourceURL
rasterToMa tch	SpatRaster	conditionally used as template raster if studyArea_rasterToMatch_biomassParam and rasterTomatch_biomassParam are not sup plied	NA
rasterToMa tch_biomassParam	SpatRaster	a raster of studyArea_biomassParam in the same resolution and projection as th e simulation's.	NA
sppColorVe ct	character	A named vector of colors to use for plot ting. The names must be in sim\$sppEquiv [[sim\$sppEquivCol]], and should also co ntain a color for 'Mixed'	NA
sppEquiv	data.table	table of species equivalencies. See LandR::sppEquivalencies_CA.	
sppNameVec tor	character	an optional vector of species names to b e pulled from sppEquiv. Species names must match P(sim)\$sppEquivCol column i n sppEquiv. If not provided, then spec ies will be taken from the entire P(sim)\$sppEquivCol column in sppEquiv. See LandR::sppEquivalencies_CA.	NA
studyArea	sf	sf polygon or terra SpatVector to use as the study area - nrow must be one	NA

objectName	objectClass	desc	sourceURL
studyArea_ biomassParam	sf	Polygon to use as the parametrisation study area. Must be provided by the user. Note that studyArea_biomassParam is only used for parameter estimation, and can be larger than the actual study area used for LandR simulations (e.g., larger than studyArea in LandR Biomass_core).	NA
studyAreaReporting	sf	multipolygon (typically smaller/unbuffered than studyArea_biomassParam and studyArea in LandR Biomass_core) to use for plotting/reporting. If not provided, will default to studyArea_biomassParam.	NA

2.2.3 List of parameters

Table 2.3 lists all parameters used in *Biomass_speciesData* and their detailed information. All these parameters have default values specified in the module's metadata.

Of these parameters, the following are particularly important:

- `coverThresh` – integer. Defines a minimum % cover value (from 0-100) that the species must have in at least one pixel to be considered present in the study area, otherwise it is excluded from the final stack of species layers (`speciesLayers`). Note that this will affect what species have data for an eventual simulation and the user will need to adjust simulation parameters accordingly (e.g., species in trait tables will need to match the species in `speciesLayers`).
- `types` – character. Which % cover data sources are to be used (see **General functioning**). Several data sources can be passed, in which case the module will overlay the lower quality layers with higher quality ones following the order of data sources in `types`. For instance, if `types == c("KNN", "CASFRI", "ForestInventory")`, KNN is assumed to be the lowest quality data set and ForestInventory the highest, hence values in KNN layers are replaced with overlapping values from CASFRI layers and values from KNN and CASFRI layers are replaced with overlapping values of ForestInventory layers.

TABLE 2.3: List of *Biomass_speciesData* parameters and their description.

paramName	paramClass	default	min	max	paramDesc	cacheKey
coverThresh	integer	10	NA	NA	The minimum % cover a species needs to have (per pixel) in the study area to be considered present	TRUE
dataYear	numeric	2020	NA	NA	Passed to <code>paste0('prepSpeciesLayers_', types)</code> function to fetch data from that year (if applicable). Defaults to 2020 as the default SCANFI year.	TRUE
sppEquivCol	character	LandR	NA	NA	The column in <code>sim\$sppEquiv</code> data.table to group species by and use as a naming convention. If different species in, e.g., the SCANFI data have the same name in the chosen column, their data are merged into one species by summing their percent cover in each raster cell.	TRUE

paramName	paramClass	default	min	max	paramDesc	cacheKey
types	character	SCANFI	NA	NA	The possible data sources. These must correspond to a function named <code>paste0('prepSpeciesLayers_', types)</code> . Defaults to 'SCANFI'. Other currently available options are: 'CASFRI', 'ForestInventory', 'KNN', 'MBFRI', 'NTEMS', 'ONFRI', 'Pickell'. All non-KNN datasets attempt to get proprietary data; the user must be granted access first. A custom function can be used to retrieve any data, just as long as it is accessible by the module (e.g., in the global environment) and is named as <code>paste0('prepSpeciesLayers_', types)</code> .	TRUE

100

2 *LandR Biomass_speciesData Module*

paramName	paramClass	default	min	max	paramDesc	cacheKey
vegLeadingProportion	numeric	0.75	0	1	a number that defines whether a species is leading for a given pixel. Only used for plotting. Default: LandR::leadingSpeciesProp(), i.e. option LandR.leadingSpeciesProp, which takes LandR.mixedwoodProp (0.75) unless set. Setting it in one place moves every module and LandR function together.	TRUE
.plotInitialTime	numeric	0	NA	NA	This describes the simulation time at which the first plot event should occur	TRUE
.plotInterval	numeric	NA	NA	NA	This describes the simulation time interval between plot events	TRUE
.plots	character	screen	NA	NA	Passed to types in Plots (see ?Plots). There are a few plots that are made within this module, if set. Note that plots (or their data) saving will ONLY occur at end(sim). If NA, plotting is turned off completely (this includes plot saving).	TRUE
.saveInitialTime	numeric	NA	NA	NA	This describes the simulation time at which the first save event should occur	TRUE
.saveInterval	numeric	NA	NA	NA	This describes the simulation time interval between save events	TRUE

paramName	paramClass	default	min	max	paramDesc	cacheKey
.sslVerify	integer	64	NA	NA	Passed to <code>httr::config(ssl_verifypeer = P(sim)\$sslVerify)</code> when downloading NFI datasets. Set to 0L if necessary to bypass checking the SSL certificate (this may be necessary when NFI's website SSL certificate is not correctly configured).	TRUE
.studyAreaName	character	NA	NA	NA	Human-readable name for the study area used. If NA, a hash of <code>studyArea_biomassParam</code> will be used.	TRUE
.useCache	character	init	NA	NA	Controls cache; caches the init event by default	TRUE
.useParallel	numeric	2	NA	NA	Used in reading csv file with <code>fread</code> . Will be passed to <code>data.table::setDTthreads</code> .	TRUE

102

2 LandR Biomass_speciesData Module

2.2.4 List of outputs

The module produces the outputs in Table 2.4, and automatically saves the processed species cover layers in the output path defined in `getPaths(sim)$outputPath`.

TABLE 2.4: List of *Biomass_speciesData* output objects and their description.

objectName	objectClass	desc
speciesLayers	SpatRaster	biomass percentage raster layers by species in Canada species map

2.2.5 Simulation flow and module events

Biomass_speciesData initialises itself and prepares all inputs provided that it has internet access to download the raw data layers, or that these layers have been previously downloaded and stored in the folder specified by `options("reproducible.destinationPath")`⁹.

The module defaults to processing cover data for all species listed in the `Boreal` column of the default `sppEquiv` input `data.table` object, for which there are available % cover layers in the `kNN` dataset (Table 2.5; see `?LandR::sppEquivalencies_CA` for more information):

TABLE 2.5: List of species cover data downloaded by default by *Biomass_speciesData*.

Species	Generic name
<i>Abies balsamea</i>	Balsam fir
<i>Abies lasiocarpa</i>	Subalpine fir
<i>Acer negundo</i>	Boxelder maple
<i>Acer pensylvanicum</i>	Striped maple
<i>Acer saccharinum</i>	Silver maple
<i>Acer saccharum</i>	Sugar maple
<i>Acer spp.</i>	Maple
<i>Acer spicatum</i>	Mountain maple

⁹Raw data layers downloaded by the module are saved in `dataPath(sim)`, which can be controlled via `options(reproducible.destinationPath = ...)`.

Species	Generic name
<i>Alnus</i> spp.	Alder
<i>Betula alleghaniensis</i>	Yellow birch
<i>Betula papyrifera</i>	Paper birch
<i>Betula populifolia</i>	Gray birch
<i>Betula</i> spp.	Birch
<i>Fagus grandifolia</i>	American beech
<i>Fraxinus americana</i>	American ash
<i>Fraxinus nigra</i>	Black ash
<i>Fraxinus</i> spp.	Ash
<i>Larix laricina</i>	Tamarack
<i>Larix lyallii</i>	Alpine larch
<i>Larix occidentalis</i>	Western larch
<i>Larix</i> spp.	Larch
<i>Picea engelmannii</i>	Engelmann's spruce
<i>Picea engelmannii</i> x <i>glauca</i>	Hybrid white x Englemann Spruce
<i>Picea glauca</i>	White spruce
<i>Picea mariana</i>	Black spruce
<i>Picea</i> spp.	Spruce
<i>Pinus albicaulis</i>	Whitebark pine
<i>Pinus banksiana</i>	Jack pine
<i>Pinus contorta</i> var. <i>latifolia</i>	Lodgepole pine
<i>Pinus monticola</i>	Western white pine
<i>Pinus resinosa</i>	Red pine
<i>Pinus</i> spp.	Pine
<i>Populus balsamifera</i>	Balsam poplar
<i>Populus balsamifera</i> v. <i>balsamifera</i>	Balsam poplar
<i>Populus grandidentata</i>	Bigtooth aspen
<i>Populus</i> spp.	Poplar
<i>Populus tremuloides</i>	Trembling aspen
<i>Populus trichocarpa</i>	Black cottonwood
<i>Tsuga canadensis</i>	Eastern hemlock
<i>Tsuga</i> spp.	Hemlock

Biomass_speciesData only runs two events, the `init` event where all species cover layers are processed and a plotting event (`initPlot`) that plots the final layers.

The general flow of *Biomass_speciesData* processes is:

1. Download (if necessary) and spatial processing of species cover layers from the first data source listed in the `types` parameter. Spatial processing consists in sub-setting the data to the area defined by `studyAreaLarge` and ensuring that the spatial projection and resolution match those of `rasterToMatchLarge`. After spatial processing, species layers that have no pixels with values $\geq$ `coverThresh` are excluded.
2. If more than one data source is listed in `types`, the second set of species cover layers is downloaded and processed as above.
3. The second set of layers is assumed to be the highest quality dataset and used to replace overlapping pixel values on the first (including for species whose layers may have been initially excluded after applying the `coverThresh` filter).
4. Steps 2 and 3 are repeated for remaining data sources listed in `types`.
5. Final layers are saved to disk and plotted (`initPlot` event). A summary of number of pixels with forest cover are calculated (`treedand numTreed` output objects; see [list of outputs](#)).

2.3 Usage example

This module can be run stand-alone, but it only compiles species % cover data into layers used by other modules.

2.3.1 Load `spades` and other packages.

2.3.2 Set up R libraries

```
options(repos = c(CRAN = "https://cloud.r-project.org"))
tempDir <- tempdir()
```

```

pkgPath <- file.path(tempDir, "packages", version$platform,
  paste0(version$major,
    ".", strsplit(version$minor, "[.]" [[1]] [1]))
dir.create(pkgPath, recursive = TRUE)
.libPaths(pkgPath, include.site = FALSE)

if (!require(Require, lib.loc = pkgPath)) {
  remotes::install_github(paste0("PredictiveEcology/",
    "Require@5c44205bf407f613f53546be652a438ef1248147"),
    upgrade = FALSE, force = TRUE)
  library(Require, lib.loc = pkgPath)
}

setLinuxBinaryRepo()

```

2.3.3 Get the module and module dependencies

```

Require(paste0("PredictiveEcology/",
  "SpaDES.project@6d7de6ee12fc967c7c60de44f1aa3b04e6eeb5db"),
  require = FALSE, upgrade = FALSE, standAlone = TRUE)

paths <- list(inputPath = normPath(file.path(tempDir,
  "inputs")),
  cachePath = normPath(file.path(tempDir, "cache")),
  modulePath = normPath(file.path(tempDir,
    "modules")), outputPath = normPath(file.path(tempDir,
    "outputs")))

SpaDES.project::getModule(modulePath = paths$modulePath,
  c("PredictiveEcology/Biomass_speciesData@master"),
  overwrite = TRUE)

## make sure all necessary packages are installed:
outs <- SpaDES.project::packagesInModules(modulePath =
  paths$modulePath)

```

```

Require(c(unname(unlist(outs)), "SpaDES"), require = FALSE,
standAlone = TRUE)

## load necessary packages
Require(c("SpaDES", "LandR", "reproducible"), upgrade = FALSE,
install = FALSE)

```

2.3.4 Setup simulation

For this demonstration we are using all default parameter values, except `coverThresh`, which is lowered to 5%. The species layers (the major output of interest) are saved automatically, so there is no need to tell `spades` what to save using the `outputs` argument (see `?SpaDES.core::outputs`).

We pass the global parameter `.plotInitialTime = 1` in the `simInitAndSpades` function to activate plotting.

```

# User may want to set some options -- see
# ?reproducibleOptions -- e.g., often the path to the
# 'inputs' folder will be set outside of project by user:
# options(reproducible.inputPaths =
# 'E:/Data/LandR_related/') # to re-use datasets across
# projects
studyAreaLarge <- Cache(randomStudyArea, size = 1e+07, cacheRepo
= paths$cachePath) # cache this so it creates a random one only
once on a machine

# Pick the species you want to work with -- here we use the
# naming convention in 'Boreal' column of
# LandR::sppEquivalencies_CA (default)
speciesNameConvention <- "Boreal"
speciesToUse <- c("Pice_Gla", "Popu_Tre", "Pinu_Con")

sppEquiv <-
LandR::sppEquivalencies_CA[get(speciesNameConvention) %in%
speciesToUse]
# Assign a colour convention for graphics for each species

```

```
sppColorVect <- LandR::sppColors(sppEquiv,
  speciesNameConvention,
  newVals = "Mixed", palette = "Set1")

## Usage example
modules <- list("Biomass_speciesData")
objects <- list(studyAreaLarge = studyAreaLarge, sppEquiv =
  sppEquiv,
  sppColorVect = sppColorVect)
params <- list(Biomass_speciesData = list(coverThresh = 5L))
```

2.3.5 Run module

Note that because this is a data module (i.e., only attempts to prepare data for the simulation) we are not iterating it and so both the start and end times are set to 1 here.

```
opts <- options(reproducible.destinationPath = paths$inputPath,
  reproducible.useCache = TRUE)

mySimOut <- simInitAndSpades(times = list(start = 1, end = 1),
  modules = modules, parameters = params, objects = objects,
  paths = paths, .plotInitialTime = 1)
options(opts)
```

Here are some of outputs of *Biomass_speciesData* (dominant species) in a randomly generated study area within Canada.

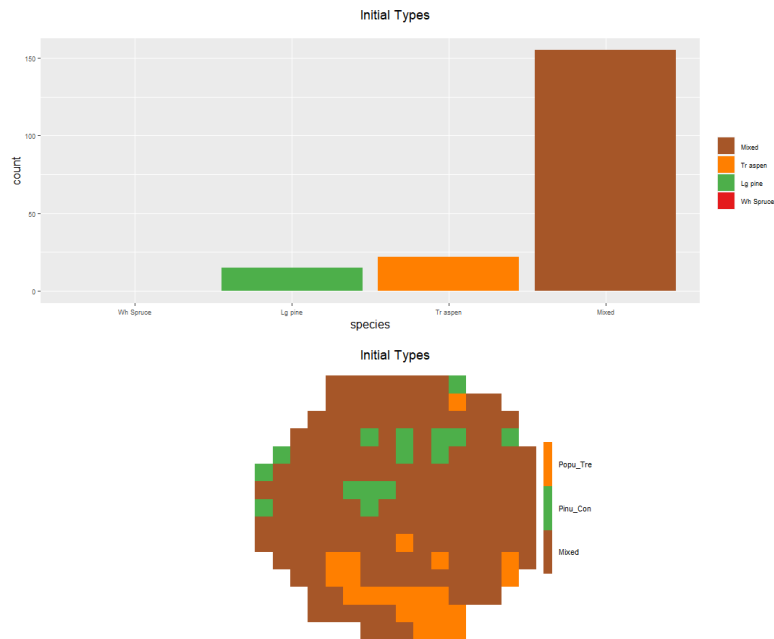


FIGURE 2.1: *Biomass_speciesData* automatically generates a plot of species dominance and number of presences in the study area when ‘.plotInitial-Time=1’ is passed as an argument.

2.4 References

- [3] A Beaudoin et al. “Mapping attributes of Canada’s forests at moderate resolution through kNN and MODIS imagery”. In: *Canadian Journal of Forest Research* 44 (2014), pp. 521–532. DOI: 10.1139/cjfr-2013-0401¹⁰.
- [4] A Beaudoin et al. *Species composition, forest properties and land cover types across Canada’s forests at 250m resolution for 2001 and 2011*. 2017. DOI: 10.23687/EC9E2659-1C29-4DDB-87A2-6ACED147A990¹¹. URL: [http :](http://)

¹⁰<https://doi.org/10.1139/cjfr-2013-0401>

¹¹<https://doi.org/10.23687/EC9E2659-1C29-4DDB-87A2-6ACED147A990>

[//open.canada.ca/data/en/dataset/ec9e2659-1c29-4ddb-87a2-6aced147a990](https://open.canada.ca/data/en/dataset/ec9e2659-1c29-4ddb-87a2-6aced147a990).

- [9] John Cosco. *COMMON ATTRIBUTE SCHEMA (CAS) FOR FOREST INVENTORIES ACROSS CANADA*. Feb. 2011, p. 117.
- [11] L Guindon et al. *SCANFI: the Spatialized CANadian National Forest Inventory data product*. 2023. DOI: 10.23687/18E6A919-53FD-41CE-B4E2-44A9707C52DC¹². URL: <https://open.canada.ca/data/en/dataset/18e6a919-53fd-41ce-b4e2-44a9707c52dc> (visited on 09/26/2025).
- [12] Luc Guindon et al. “A new approach for spatializing the Canadian National Forest Inventory (SCANFI) using Landsat dense time series”. en. In: *Canadian Journal of Forest Research* 54.7 (July 2024), pp. 793–815. ISSN: 0045-5067, 1208-6037. DOI: 10.1139/cjfr-2023-0118¹³. URL: <https://cdnsiencepub.com/doi/10.1139/cjfr-2023-0118> (visited on 09/26/2025).
- [13] Txomin Hermosilla et al. “Characterizing long-term tree species dynamics in Canada’s forested ecosystems using annual time series remote sensing data”. en. In: *Forest Ecology and Management* 572 (Nov. 2024), p. 122313. ISSN: 03781127. DOI: 10.1016/j.foreco.2024.122313¹⁴. URL: <https://linkinghub.elsevier.com/retrieve/pii/S037811272400625X> (visited on 09/26/2025).
- [22] Paul D Pickell and Nicholas C Coops. *Development of historical forest attribute layers using Landsat time series and kNN imputation for the western Canadian boreal forest*. Tech. rep. University of British Columbia, Dec. 2016, pp. 1–27.

¹²<https://doi.org/10.23687/18E6A919-53FD-41CE-B4E2-44A9707C52DC>

¹³<https://doi.org/10.1139/cjfr-2023-0118>

¹⁴<https://doi.org/10.1016/j.foreco.2024.122313>

3

LandR Biomass_borealDataPrep Module

Biomass borealDataPrep v. 1.8.0.9005 ¹

Get help Report issues ²

3.0.0.1 Authors:

Yong Luo Yong.Luo@gov.bc.ca³ [aut], Eliot J B McIntire eliot.mcintire@nrcan-rncan.gc.ca⁴ [aut, cre], Ceres Barros ceres.barros@ubc.ca⁵ [aut], Alex M. Chubaty achubaty@for-cast.ca⁶ [aut]

This documentation is a work in progress. Potential discrepancies and omissions may exist for the time being. If you find any, contact us using the “Get help” link above.

3.1 Module Overview

3.1.1 Quick links

- General functioning
- List of input objects
- List of parameters

¹https://github.com/PredictiveEcology/Biomass_borealDataPrep15e3151fa06f4cd9f2ccf5bf75d0310fda1ac33c

²https://github.com/PredictiveEcology/Biomass_borealDataPrep/issues

³<mailto:Yong.Luo@gov.bc.ca>

⁴<mailto:eliot.mcintire@nrcan-rncan.gc.ca>

⁵<mailto:ceres.barros@ubc.ca>

⁶<mailto:achubaty@for-cast.ca>

- **List of outputs**
- **Simulation flow and module events**

3.1.2 Summary

LandR *Biomass_borealDataPrep* (hereafter *Biomass_borealDataPrep*), prepares all necessary inputs for *Biomass_core* based on data available for forests across Canada forests, but focused on Western Canada boreal forest systems. Nevertheless, it provides a good foundation to develop other modules aimed at different geographical contexts. By keeping data preparation and parameter estimation outside of *Biomass_core*, we promote the modularity of the LandR-based model systems and facilitate interoperability with other parameter estimation procedures.

Specifically, it prepares and adjusts invariant and spatially-varying species trait values, as well as ecolocation-specific parameters, probabilities of germination and initial conditions necessary to run *Biomass_core*. For this, *Biomass_borealDataPrep* requires internet access to retrieve default data⁷.

We advise future users to run *Biomass_borealDataPrep* with defaults and inspect the resulting input objects are like before supplying alternative data (or data URLs).

3.1.3 Links to other modules

Biomass_borealDataPrep is intended to be used with *Biomass_core*⁸, but can be linked with other data modules that prepare inputs. See here⁹ for all available modules in the LandR ecosystem and select *Biomass_borealDataPrep* from the drop-down menu to see potential linkages.

- *Biomass_core*¹⁰: core forest dynamics simulation module. Used downstream from *Biomass_borealDataPrep*;
- *Biomass_speciesData*¹¹: grabs and merges several sources of species cover

⁷Raw data layers downloaded by the module are saved in `dataPath(sim)`, which can be controlled via `options(reproducible.destinationPath = ...)`.

⁸https://github.com/PredictiveEcology/Biomass_core

⁹https://rpubs.com/PredictiveEcology/LandR_Module_Ecosystem

¹⁰https://github.com/PredictiveEcology/Biomass_core

¹¹https://github.com/PredictiveEcology/Biomass_speciesData

data, making species percent cover (% cover) layers used by other LandR Biomass modules. Default source data spans the entire Canadian territory. Used upstream from *Biomass_borealDataPrep*;

- *Biomass_speciesParameters*¹²: calibrates four-species level traits using permanent sample plot data (i.e., repeated tree biomass measurements) across Western Canada. Used downstream from *Biomass_borealDataPrep*.

3.2 Module manual

3.2.1 General functioning

Biomass_borealDataPrep prepares all inputs necessary to run a realistic simulation of forest dynamics in Western Canadian boreal forests using *Biomass_core*. Part of this process involves cleaning up the input data and imputing missing data in some cases, which are discussed in detail in [Data acquisition and treatment](#).

After cleaning and formatting the raw input data, the module:

1. **calculates species biomass per pixel** by multiplying the observed species % cover by the observed stand biomass and an adjustment factor, which can be statistically calibrated for the study area. Given that this adjusts the species biomass, this calibration step contributes to the calibration of `maxB` and `maxANPP` trait values, whose estimation is also based on species biomass (see [Initial species age and biomass per pixel](#) and [Adjustment of species biomass](#));
2. prepares **invariant species traits** – these are spatio-temporally constant species traits that influence population dynamics (e.g., growth, mortality, dispersal) and responses to fire (see [Invariant species traits](#));
3. defines **ecolocations** – groupings of pixels with similar biophysical conditions. By default, ecolocations are defined as the spatial combination of ecodistricts of the National Ecological Framework

¹²https://github.com/PredictiveEcology/Biomass_speciesParameters

for Canada, and the Land Cover of Canada 2010 map (see [Defining simulation pixels and ecolocations](#)). **Note that ecolocation is termed `ecoregionGroup` across LandR modules.**

4. prepares **ecolocation-specific parameters** and **probabilities of germination** – only one ecolocation-specific parameter is used, the **minimum relative biomass thresholds**, which defines the level of shade in a pixel. Together the level of shade and the **probabilities of germination** influence germination success in any given pixel;
5. estimates **spatio-temporally varying species traits** – species traits that can vary by ecolocation and in time. These are maximum biomass (`maxB`), maximum above-ground net primary productivity (`maxANPP`; see [Maximum biomass and maximum aboveground net primary productivity](#)) and species establishment probability (`SEP`, called `establishprob` in the module traits table; see [Species establishment probability](#)). **By default, *Biomass_borealDataPrep* estimates temporally constant values of `maxB`, `maxANPP` and `SEP`;**
6. creates **initial landscape conditions** – *Biomass_borealDataPrep* performs data-based landscape initialisation, by creating the species cohort table (`cohortData`) and corresponding map (`pixelGroupMap`; both used to initialise and track cohorts across the landscape) based on observed stand age and species biomass.

As *Biomass_core* only simulates tree species dynamics, *Biomass_borealDataPrep* prepares all inputs and estimates parameters in pixels within forested land-cover classes (see [Defining simulation pixels and ecolocations](#)).

If a `studyArea_biomassParam` is supplied, the module uses it for parameter estimation to account for larger spatial variability.

In the next sections, we describe in greater detail the various data processing and parameter estimation steps carried out by *Biomass_borealDataPrep*.

3.2.2 Data acquisition and treatment

The only two objects that the user must supply are spatial polygons that define the study area used to derive parameters (`studyArea_biomassParam`), and the study area where the simulation will happen (`studyArea`). The two objects can be identical if the user chooses to parametrise and run the simulations in the same area. If not identical, `studyArea` must be fully within `studyArea_biomassParam`. If `studyArea_biomassParam` and `studyArea` are

in Canada, the module can automatically estimate and prepare all input parameters and objects for *Biomass_core*. Best attempts are made to use FAIR data [sensu 32] at the national-scale; however, the default SCANFI 2020 datasets are unpublished, and not yet publicly available. The previous default using the Canadian National Forest Inventory (NFI) forest attributes kNN dataset [kNN; Beaudoin et al. [3], Beaudoin et al. [4]] for `dataYear = 2001` or `2011` is still available.

If no other inputs are supplied, *Biomass_borealDataPrep* will create raster layer versions `studyArea_biomassParam` and `studyArea(rasterToMatch_biomassParam and rasterToMatch`.

3.2.2.1 Defining simulation pixels and ecolocations

Biomass_borealDataPrep uses land-cover data to define and assign parameter values to the pixels where forest dynamics will be simulated (forested pixels). By default it uses land-cover classes derived from the Spatialized Canadian National Forest Inventory [SCANFI; Guindon et al. [11], Guindon et al. [12]], a raster-based database that distinguishes several forest and non-forest land-cover types. The original SCANFI LCC codes are substituted with their corresponding values according to the National Terrestrial Ecosystem Monitoring System for Canada Land Cover [NTEMS; Hermosilla et al. [14]] classification. Pixels with classes 81, 210, 220, 230, and 240 are included as forested pixels (see parameter `forestedLCCClasses`).

When the land-cover raster (`rstLCC`) includes transient cover types (e.g., recently disturbed) the user may pass a vector of transient class IDs (via the parameter `LCCClassesToReplaceNN`) that will be reclassified into a “stable” forested class (defined via the parameter `forestedLCCClasses`). The reclassification is done by searching the focal neighbourhood for a replacement forested cover class (up to a radius of 1250m from the focal cell). If no forested class is found within this perimeter, the pixel is not used to simulate forest dynamics. Reclassified pixels are omitted from the fitting of statistical models used for parameter estimation, but are assigned predicted values from these models.

Sub-regional spatial variation in `maxBiomass`, `maxANPP`, and SEP species traits is accounted for by ecolocation. Ecolocations are used as proxies for biophysical variation across the landscape when estimating model parameters that vary spatially. By default, they are defined as the combination of “ecodistricts” from the National Ecological Framework for

Canada¹³ (a broad-scale polygon layer that captures sub-regional variation) (`ecoregionLayer`) and the above land cover (`rstLCC`), but the user can change this by supplying different ecozonation or land-cover layers.

3.2.2.2 Species cover

Species cover (% cover) raster layers (`speciesLayers`) can be automatically obtained and pre-processed by *Biomass_borealDataPrep*. The module ensures that:

1. all data have the same geospatial properties (extent, resolution);
2. all layers these are correctly re-projected to `studyArea_biomassParam` and `rasterToMatch_biomassParam`;
3. species with no cover values above 10% are excluded.

By default it uses species % cover rasters derived from the Spatialized Canadian National Forest Inventory [SCANFI; Guindon et al. [11], Guindon et al. [12]] – hereafter ‘SCANFI species data’.

3.2.2.3 Initial species age and biomass per pixel

Stand age and aboveground stand biomass (hereafter ‘stand biomass’) are used to derive parameters and define initial species age and biomass across the landscape. These are also derived from SCANFI [11, @Guindon2024] by default.

Biomass_borealDataPrep downloads these data and performs a number of data harmonization operations to deal with data inconsistencies. It first searches for mismatches between stand age (`standAge`), stand biomass (`standB`) and total stand cover (`standCover`), assuming that cover is the most accurate of the three, and biomass the least, and in the following order:

1. Pixels with `standCover < 5%` are removed;
2. Pixels with `standAge == 0`, are assigned `standB == 0`;
3. Pixels with `standB == 0`, are assigned `standAge == 0`.

Then, species is assigned one cohort per pixel according to the corrected stand age, stand biomass and % cover values. Cohort age is assumed to be the

¹³<https://open.canada.ca/data/en/dataset/3ef8e8a9-8d05-4fea-a8bf-7f5023d2b6e1>

same as stand age and biomass is the product of stand biomass and species % cover. Before doing so, stand cover is rescaled to vary between 0 and 100%.

A next set of data inconsistencies in cohort age (age), biomass (B) and cover (cover) is looked for and solved in the following order:

4. if `cover > 0` and `age == 0`, B is set to 0 (and stand biomass recalculated);
5. if `cover == 0` and `age > 0`, or if `age == NA`, age is empirically estimated using the remainder of the data to fit the model supplied by `P(sim)$imputeBadAgeModel`, which defaults to:

```
## [[1]]
## lme4::lmer(log(age) ~ log(totalBiomass) * cover * speciesCode +
##           (log(totalBiomass) | initialEcoregionCode))
```

Cohort biomass is then adjusted to reflect the different cover to biomass relationship of conifer and broadleaf species (see [Adjustment of initial species biomass](#)).

3.2.2.4 Replacing initial biomass and age within known fire perimeters

Biomass_borealDataPrep can use fire perimeters to correct stand ages. To do so, it downloads the latest fire perimeter data from the Canadian Wildfire Data Base¹⁴ and changes pixel age inside fire perimeters to match the time since last fire, using fire years up to the first year of the simulation.

Taking two independent datasets for stand age (fire perimeters) and stand biomass (derived from MODIS satellite imagery) can cause discrepancies (e.g., stand age = 5 and aboveground biomass = 10000 m^2/ha). This may be due to errors coming from: a) a stand replacing disturbance that reset age to zero a few years before, but the biomass layer was not zeroed; b) the disturbance was not stand-replacing (leaving biomass), but age was still zeroed. This means that either, aboveground biomass is wrong or age is.

Options to address this include 1) get better data for these two variables that do not contradict one another (not currently available to us) or 2) estimate one or the other. There is no obvious way to decide which one is incorrect, unless there is an independent data source.

¹⁴<https://cwffis.cfs.nrcan.gc.ca/datamart>

In the current *Biomass_borealDataPrep* module version, we chose to correct both. If `P(sim)$fireURL` is provided and `P(sim)$overrideBiomassInFires` is TRUE, fire perimeters are used as the source of information for age, and *Biomass_core* is used to generate corresponding biomass values based on the estimated growth parameters and known species presence/absence (from the species cover layers).

This assumes that 1) recorded fires were stand-replacing, and so time since fire is the new stand age and 2) that the first year of the simulation is later than the first fire year in the fire perimeter data. The biomass spin-up with *Biomass_core* is only run in pixels where stand ages were corrected, for as long as the new stand age (i.e., the time since last fire). All species start with age = 0 and biomass = 0, and grow until time since last fire is achieved. The resulting species biomass is used as the initial biomass values for each species cohort in the actual simulation.

If the user does not want to perform this imputation, this step can be bypassed by setting the parameter `P(sim)$overrideBiomassInFires` to FALSE or `P(sim)$fireURL` to NULL or NA.

Also pixels that suffered *any kind* of data imputation (e.g., the age corrections detailed in the [previous section](#)) can be excluded from the simulation by setting `P(sim)$rmImputedPix == TRUE`.

3.2.2.5 Invariant species traits

Most invariant species traits are obtained from available species trait tables used in LANDIS-II applications in Canada's boreal forests (available in Dominic Cyr's GitHub repository¹⁵). Some are then adapted with minor adjustments to match Western Canadian boreal forests using published literature. Others (key growth and mortality traits) can be calibrated by *Biomass_speciesParameters* (see [Calibrating species growth/mortality traits using Biomass_speciesParameters](#)).

The LANDIS-II species trait table contains species trait values for each Canadian Ecozone [20], which are by default filtered to the Boreal Shield West (BSW), Boreal Plains (BP) and Montane Cordillera Canadian Ecozones (via `P(sim)$speciesTableAreas`). Most trait values do not vary across these ecozones for a given species, but when they do the minimum value is used.

The function `LandR::speciesTableUpdate` is used by default to do further

¹⁵https://github.com/dcyr/LANDIS-II_IA_generalUseFiles

adjustments to trait values in this table (if this is not intended, a custom function call or NULL can be passed to `P(sim)$speciesUpdateFunction`):

- Longevity values are adjusted to match the values from Burton and Cumming [6], which match BSP, BP and MC ecozones. These adjustments result in higher longevity for most species;
- Shade tolerance values are lowered for *Abies balsamifera*, *Abies lasiocarpa*, *Picea engelmannii*, *Picea glauca*, *Picea mariana*, *Tsuga heterophylla* and *Tsuga mertensiana* to better **relative** shade tolerance levels in Western Canada. Because these are relative shade tolerances, the user should **always** check these values with respect to their own study areas and species pool.

The user can also pass more than one function call to `P(sim)$speciesUpdateFunction` if they want to make other adjustments in addition to those listed above (see `?LandR::updateSpeciesTable`).

3.2.2.6 Probabilities of germination

By default, *Biomass_borealDataPrep* uses the same probabilities of germination (called `sufficientLight` in the module) as *Biomass_core*. These are obtained from publicly available LANDIS-II table¹⁶.

3.2.3 Parameter estimation/calibration

3.2.3.1 Adjustment of initial species biomass

Biomass_borealDataPrep estimates initial values of species aboveground biomass (B) based on stand biomass (`standB`) and individual species % cover. Initial B is estimated for each species in each pixel by multiplying `standB` by species % cover. Because the default cover layers are satellite-derived, the relationship between relative cover and relative biomass of broadleaf and conifer species needs to be adjusted to reflect their different canopy architectures (using `P(sim)$deciduousCoverWeight`).

`P(sim)$deciduousCoverWeight` is how much biomass a unit of deciduous cover carries relative to a unit of conifer cover. Broadleaf crowns are wider per unit of wood, so it is expected below 1.

¹⁶https://raw.githubusercontent.com/LANDIS-II-Foundation/Extensions-Succession/master/biomass-succession-archive/trunk/tests/v6.0-2.0/biomass-succession_test.txt

By default (`P(sim)$fitDeciduousCoverWeight == TRUE`) it is **estimated from studyArea_biomassParam** rather than taken from the parameter, because it is not a universal constant: on 100 km of Alberta boreal mixedwood the fit gives 0.90, and per ecoregion inside that one window it ranges 0.80 to 1.11.

`partitionBiomass()` splits a pixel's `standB` among its cohorts and so says nothing about the total it divides; what identifies the weight is a statement about the total,

$$standB_p = k_p \times (C_p + w \times D_p)$$

where C_p and D_p are the pixel's total conifer and deciduous cover. What k decides what is estimated. Held constant within an ecoregion, w absorbs site quality – in the boreal, broadleaf grows on richer, better-drained ground than black spruce, so the estimate comes back **above 1** and has credited the site to the species. Made a function of the pixel's own canopy height and closure (`sim$rstCanopyHeight`, `sim$rstCanopyClosure`, SCANFI's own layers), it compares composition between stands carrying the *same* amount of structure, which is the canopy-architecture difference the parameter is for. That is what the module fits, by profiling w with k concentrated out; the design matrix does not depend on w , so the whole search costs a few seconds.

Where the study area cannot identify it – too little deciduous cover, no canopy height and closure layers (only SCANFI publishes them), or an estimate sitting on the search bound – the module says so and keeps `P(sim)$deciduousCoverWeight`. Set `P(sim)$fitDeciduousCoverWeight == FALSE` to always use the parameter.

3.2.3.2 Maximum biomass and maximum aboveground net primary productivity

Biomass_borealDataPrep statistically estimates maximum biomass (`maxB`) and maximum aboveground net primary productivity (`maxANPP`) using the processed species ages and biomass.

`maxB` is estimated by modelling the response of species biomass (`B`) to species age and cover, while accounting for variation between ecolocations (`ecoregionGroup` below):

```
## [[1]]
## lme4::lmer(B ~ logAge * speciesCode + cover * speciesCode + (logAge +
##          cover | ecoregionGroup))
```

The coefficients are estimated by maximum likelihood and model fit is calculated as the proportion of explained variance explained by fixed effects only (marginal r^2) and by the entire model (conditional r^2) – both of which are printed as messages.

Because the model can take a while to fit, by default we sample pixels within each species and ecolocation combination (sample size defined by the `P(sim)$subsetDataBiomassModel` parameter).

If convergence issues occur and `P(sim)$fixModelBiomass == TRUE`, the module attempts to refit the model by re-sampling the data, re-fitting `lmer` with the `bobyqa` optimizer, and re-scaling the continuous predictors (by default, `cover` and `logAge`). These steps are tried additively until the convergence issue is resolved. If the module is still unable to solve the convergence issue a message is printed and the module uses the last fitted model.

Note that convergence issues are not usually problematic for the estimation of coefficient values, only for estimation of their standard errors. However, the user should always inspect the final model (especially if not converged) and make sure that the problems are not significant and that the fitted model meets residual assumptions. For this, the user should make sure model objects are exported to the `simList` using the `exportModels` parameter.

Alternative model calls/formulas can be supplied via the `P(sim)$biomassModel` parameter. Note that if supplying a model call that does not use `lme4::lmer` the refitting process is likely to fail and may have to be disabled (via the `P(sim)$fixModelBiomass` parameter).

Another consideration to add with respect to the estimation of `maxB`, is that we are choosing a linear model to relate $B \sim \log(\text{age}) + \text{cover}$. This is not ideal from an ecological point of view, as biomass is unlikely to vary linearly with age or cover, and more likely to saturate beyond a certain high value of cover and follow a hump-shaped curve with age (i.e., reaching maximum values for a given age, and then starting to decrease as trees approach longevity). Also, fitting a linear model can lead to negative B values at young ages and low cover. However, our tests revealed that a linear mixed effects model was not producing abnormal estimates of B at maximum values of age and cover (hence, `maxB` estimates), while allowing to leverage on the powerful statistical machinery of `lme4`.

Finally, we highlight that modelling $\log(B)$ is NOT an appropriate solution, because it will wrongly assume an *exponential* relationship between $B \sim \log(\text{age}) + \text{cover}$, leading to a serious overestimation of `maxB` (Fig. 3.1) and

steep increases in species biomasses during the first years of the simulation (Fig. 3.2).

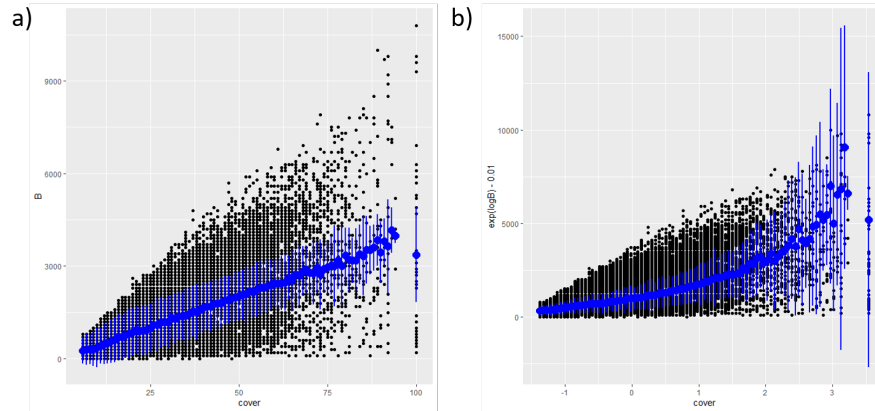


FIGURE 3.1: Modelling biomass as a linear vs. exponential relationship. a) ‘modelBiomass’ as ‘ $B \sim \log \text{Age} * \text{speciesCode} + \text{cover} * \text{speciesCode} + (\log \text{Age} + \text{cover} \mid \text{ecoregionGroup})$ ’. b) ‘modelBiomass’ as ‘ $\log B \sim \log \text{Age} * \text{speciesCode} + \text{cover} * \text{speciesCode} + (\log \text{Age} + \text{cover} \mid \text{ecoregionGroup})$ ’. Blue dots are marginal mean B values (back-transformed in b) cross ages with confidence intervals as the bars.

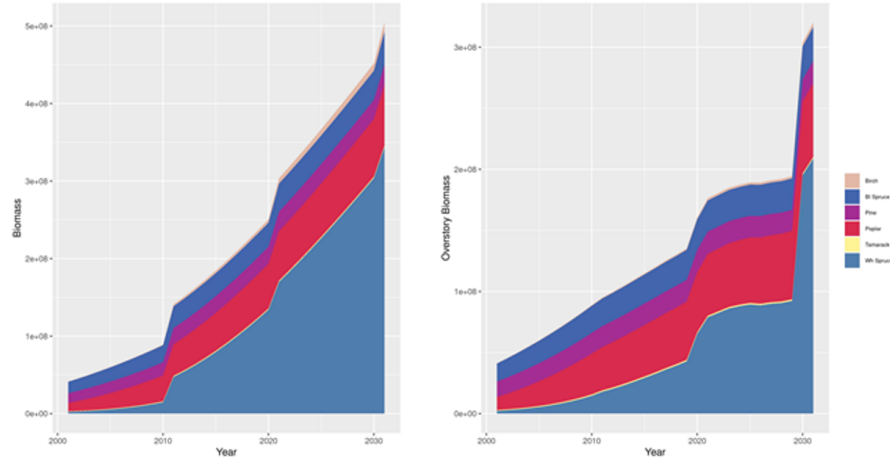


FIGURE 3.2: Thirty years of simulation with ‘maxB’ values estimated from a ‘logB ...’ ‘biomassModel’ (see Fig. reffig:fig-biomassModelLogBtest). The steep increase in such little time is abnormal.

After the biomass model is fit, maxB is predicted by species and ecolocation combination, for maximum species cover values (100%) and maximum log-age (the log of species longevity). When using *Biomass_speciesParameters*, maxB is calibrated so that species can achieve the maximum observed biomass during the simulation (see [Calibrating species growth/mortality traits using *Biomass_speciesParameters*](#)).

maxANPP is calculated as $\text{maxB} * \text{mANPPproportion}/100$, where mANPPproportion defaults to 3.33, unless calibrated by *Biomass_speciesParameters* (see [Calibrating species growth/mortality traits using *Biomass_speciesParameters*](#)). The default value, 3.33, comes from an inversion of the rationale used to calculate maxB in Scheller and Mladenoff [27]. There, the authors estimated maxANPP using the model PnET-II (and then adjusted the values manually) and from these estimates calculated maxB by multiplying the estimated maxANPP by 30.

3.2.3.3 Species establishment probability

Species establishment probability (SEP, `establishprob` in the module) is estimated by modelling the probability of observing a given species in each ecolocation. For this, *Biomass_borealDataPrep* models the relationship be-

tween probability of occurrence of a species (π) using the following model by default:

```
## [[1]]
## glm(cbind(coverPres, coverNum - coverPres) ~ speciesCode * ecoregionGroup,
##      family = binomial)
```

whereby the probability of occurrence of a species (π) – calculated as the proportion of pixels with % cover > 0 – is modelled per species and ecolocation following a binomial distribution with a logit link function. There is no data sub-sampling done before fitting the SEP statistical model, as the model fits quickly even for very large sample sizes (e.g., > 20 million points).

SEP is then predicted by species and ecolocation combination and the resulting values are integrated over the length of the succession time step (successionTimestep parameter) as:

$$integratedSEP = 1 - (1 - estimatedSEP)^{e^{successionTimestep}} \quad (3.1)$$

This is important, since seed establishment only occurs once at every $P(sim)$successionTimestep$, and thus the probabilities of seed establishment need to be temporally integrated to reflect the probability of a seed establishing in this period of time.

Finally, since the *observed* species cover used to fit `coverModel` is a result of both seed establishment and resprouting/clonal growth, the final species establishment probabilities are calculated as a function of the temporally integrated presence probabilities and species' probabilities of resprouting (`resproutprob`, in the species table) (bounded between 0 and 1):

$$SEP = integratedSEP * (1 - resproutprob) \quad (3.2)$$

if $SEP > 1$, then:

$$SEP = 1 \quad (3.3)$$

if $SEP < 0$, then:

$$SEP = 0 \quad (3.4)$$

3.2.3.4 Ecolocation-specific parameter – minimum relative biomass

Minimum relative biomass (`minRelativeB`) is a spatially-varying parameter used to determine the shade level in each pixel. Each shade class (X0-X5) is defined by a minimum relative biomass threshold compared to the pixel's current relative biomass, which is calculated as the sum of pixel's total biomass divided by the total potential maximum biomass in that pixel (the sum of all `maxB` for the pixel's ecolocation).

Since we found no data to base the parametrisation of the shade class thresholds, default values are based on publicly available values used in LANDIS-II applications in Canada's boreal forests (available in Dominic Cyr's GitHub repository¹⁷), and all ecolocations share the same values.

Initial runs revealed excessive recruitment of moderately shade-intolerant species even as stand biomass increased, so values for shade levels X4 and X5 are adjusted downwards (X4: 0.8 to 0.75; X5: 0.90 to 0.85) to reflect higher competition for resources (e.g. higher water limitation) in Western Canadian forests with regards to Eastern Canadian forests [19], which are likely driven by higher moisture limitation in the west [15, 21].

This adjustment can be bypassed by either supplying a `minRelativeB` table, or an alternative function call to `P(sim)$minRelativeBFunction` (which by default is `LandR::makeMinRelativeB`).

The minimum biomass threshold of a shade level of X0 is 0 `standB`.

3.2.3.5 Calibrating species growth/mortality traits using *Biomass_speciesParameters*

If using *Biomass_borealDataPrep* and *Biomass_speciesParameters*, the later module calibrates several species traits that are first prepared by *Biomass_borealDataPrep*:

- `growthcurve`, `mortalityshape` – which initially come from publicly available LANDIS-II tables;
- `maxBiomass`, `maxANPP` – which are estimated statistically by *Biomass_borealDataPrep* (see **Maximum biomass and maximum above-ground net primary productivity**).

¹⁷https://github.com/dcy/LANDIS-II_IA_generalUseFiles

Briefly, *Biomass_speciesParameters*:

1. Uses ~41,000,000 hypothetical species' growth curves (generated with *Biomass_core*), that cover a fully factorial combination of longevity, ratio of maxANPP to maxBiomass, growthcurve, mortalityshape;
2. Takes permanent and temporary sample plot (PSP) data in or near the study area for the target species, and finds which hypothetical species' growth curve most closely matches the growth curve observed in the PSP data – on a species-by-species base. This gives us each species' growthcurve, mortalityshape, and a new species trait, *manPPproportion*, a ratio of maximum aboveground net primary productivity (maxANPP) to maximum biomass (maxBiomass, not to be confounded with maxB) in the study area.
3. Introduces a second new species trait, *inflationFactor*, and recalibrates maxB. We recognize that maxB, as obtained empirically by *Biomass_borealDataPrep*, cannot be easily reached in simulations because all reasonable values of growthcurve, mortalityshape and longevity prevent the equation from reaching maxB (it acts as an asymptote that is never approached). The *inflationFactor* is calculated as the ratio of maxBiomass (the parameter used to generate theoretical growth curves in step 1) to the maximum biomass *actually* achieved by the theoretical growth curves (step 1). maxB is then recalibrated by multiplying it by *inflationFactor*. By doing this, resulting non-linear growth curves generated doing *Biomass_core* simulation will be able to achieve the the empirically estimated maxB.
4. Estimates species-specific maxANPP by multiplying the final maxB above by *manPPproportion* (estimated in step 2).

In cases where there is insufficient PSP data to perform the above steps, maxB and maxANPP are left as estimated by *Biomass_borealDataPrep* (see **Maximum biomass and maximum aboveground net primary productivity**) and *inflationFactor* and *manPPproportion* take default values of 1 and 3.33.

3.2.4 Aggregating species

Biomass_borealDataPrep will use the input table *sppEquiv* and the parameter *P(sim)\$sppEquivCol* to select the naming convention to use for the sim-

ulation (see full list of **input objects** and **parameters** for details). The user can use this table and parameter to define groupings that “merge” similar species that have their own invariant trait values (see **Invariant species traits**) (e.g. genus-level group or a functional group). To do so, the name of the species group in sppEquivCol column of the sppEquiv table must be identical for each grouped species.

TABLE 3.1: Example of species merging for simulation. Here the user wants to model *Abies balsamea*, *A. lasiocarpa* and *Pinus contorta* as separate species, but all *Picea spp.* as a genus-level group. For this, all six species are identified in the ‘SCANFI’ column, so that their % cover layers can be obtained, but in the ‘Boreal’ column (which defines the naming convention used in the simulation in this example) all *Picea spp.* have the same name. *Biomass_borealDataPrep* will merge their % cover data into a single layer by summing their cover per pixel. (continued below)

Species	SCANFI	Boreal
<i>Abies balsamea</i>	ABIE_BAL	Abie_Bal
<i>Abies lasiocarpa</i>	ABIE_LAS	Abie_Las
<i>Picea engelmannii</i>	PICE_ENG	Pice_Eng
<i>Picea engelmannii x glauca</i>	PICE_ENG_GLA	Pice_Eng_Gla
<i>Picea glauca</i>	PICE_GLA	Pice_Gla
<i>Picea mariana</i>	PICE_MAR	Pice_Mar
<i>Picea spp.</i>		Pice_Spp
<i>Pinus contorta</i>		Pinu_Con
<i>Pinus contorta</i> var. <i>contorta</i>	PINU_CON_CON	Pinu_Con
<i>Pinus contorta</i> var. <i>latifolia</i>	PINU_CON_LAT	Pinu_Con

Modelled as
<i>Abies balsamea</i>
<i>Abies lasiocarpa</i>
<i>Picea engelmannii</i>
<i>Picea engelmannii x glauca</i>
<i>Picea glauca</i>
<i>Picea mariana</i>
<i>Picea spp.</i>
<i>Pinus contorta</i>
<i>Pinus contorta</i> var. <i>contorta</i>

 Modelled as

Pinus contorta var. latifolia

When groups contain species with different (invariant) trait values, the minimum value across all species is used. As for the default species % cover layers, *Biomass_borealDataPrep* proceeds in the same way as *Biomass_speciesData* and sums cover across species of the same group per pixel.

3.2.5 List of input objects

Below are the full lists of input objects (Table 3.2) that *Biomass_borealDataPrep* expects.

The only inputs that **must** be provided (i.e., *Biomass_borealDataPrep* does not have a default for) are *studyArea* (the study area used to simulate forest dynamics *Biomass_core*) and *studyArea_biomassParam* (a potentially larger study area used to derive parameter values – e.g., species traits).

All other input objects and parameters have internal defaults.

Of these inputs, the following are particularly important and deserve special attention:

Spatial layers

- *ecoregionLayer* or *ecoregionRst* – a shapefile or map containing ecological zones.
- *rawBiomassMap* – a map of observed stand biomass (in g/m^2).
- *rstLCC* – a land-cover raster.
- *speciesLayers* – layers of species % cover data. The species must match those available in default (or provided) species traits tables (the *species* and *speciesEcoregion* tables).
- *standAgeMap* – a map of observed stand ages (in years).
- *studyArea* – shapefile. A *SpatialPolygonsDataFrame* with a single polygon determining where the simulation will take place. This input object **must be supplied by the user**.

- `studyArea_biomassParam` – `shapefile`. A `SpatialPolygonsDataFrame` with a single polygon determining the where the statistical models for parameter estimation will be fitted. It **must** contain `studyArea` fully, if they are not identical. This object **must be supplied by the user**.

Tables

- `speciesTable` – a table of invariant species traits that must have the following columns (even if not all are necessary to the simulation): “species”, “Area”, “longevity”, “sexualmature”, “shadetolerance”, “firetolerance”, “seeddistance_eff”, “seeddistance_max”, “resproutprob”, “resproutage_min”, “resproutage_max”, “postfireregen”, “leaflongevity”, “wooddecayrate”, “mortalityshape”, “growthcurve”, “leafLignin”, “hardsoft”. The order can vary but the column names must be identical. See Scheller and Miranda [26] and *Biomass_core* manual for further detail about these columns.

TABLE 3.2: List of *Biomass_borealDataPrep* input objects and their description.

objectName	objectClass	desc	sourceURL
cloudFolderID	character	The google drive location where cloudCache will store large statistical objects	NA
columnsForPixelGroups	character	The names of the columns in cohortData that define unique pixelGroups. Default is c('ecoregionGroup', 'speciesCode', 'age', 'B'); see ?LandR::columnsForPixelGroups().	NA
ecoregionLayer	sf	A sf polygon object that characterizes the unique ecological regions (ecoregionGroup) used to parameterize the biomass, cover, and species establishment probability models. It will be overlaid with landcover to generate classes for every ecoregion/LCC combination. It must have same extent and crs as studyArea_biomassParam. It is superseded by sim\$ecoregionRst if that object is supplied by the user	https://sis.agr.gc.ca/cansis/nadb/ecostat/district/ecodistrict_shp.zip

objectName	objectClass	desc	sourceURL
ecoregionRst	SpatRaster	A raster that characterizes the unique ecological regions used to parameterize the biomass, cover, and species establishment probability models. If this object is provided, it will supercede <code>sim\$ecoregionLayer</code> . It will be overlaid with <code>landcover</code> to generate classes for every ecoregion/LCC combination. It must have same extent and crs as <code>rasterToMatch_biomassParam</code> if supplied by user - use <code>reproducible::postProcess</code> . If it uses an attribute table, it must contain the field 'ecoregion' to represent raster values	NA
firePerimeters	SpatRaster	Fire perimeters raster, with fire year information used to 'update' stand age using time since last fire as the imputed value. Only used if <code>P(sim)\$overrideAgeInFires = TRUE</code> . Biomass will also be updated in these pixels if <code>P(sim)\$overrideBiomassInFires = TRUE</code> and the last fire was later than 1985. Defaults to using fire perimeters in the Canadian National Fire Database, downloaded as a zipped shapefile with fire polygons, an attribute (i.e., a column) named 'YEAR', which is used to rasterize to the study area.	https://cwfis.cfs.nrcan.gc.ca/downloads/nfdb/fire_poly/current_version/NFDB_poly.zip

objectName	objectClass	desc	sourceURL
imputedPix ID	integer	A vector of pixel IDs - matching rasterM atch IDs - that suffered data imputation . Data imputation may be in age (to matc h last fire event post 1950s, or O cover), biomass (to match fire-related impute d ages; correct for missing values or fo r O age/cover), land cover (to convert n on-forested classes into to nearest fore sted class). If standAgeMap had impute d data, then this is expected to be crea ted at that time. It will be added as an attribute to sim\$standAgeMap	NA

objectName	objectClass	desc	sourceURL
rstLCC	SpatRaster	<p>A land classification map in study area. It must be 'corrected', in the sense that: 1) Every class must not conflict with any other map in this module (e.g., speciesLayers should not have data in LCC classes that are non-treed); 2) It can have treed and non-treed classes. The non-treed will be removed within this module if <code>P(sim)\$omitNonTreedPixels</code> is TRUE; 3) It can have transient pixels, such as 'young fire'. These will be converted to the nearest non-transient class, probabilistically if there is more than 1 nearest neighbour class, based on <code>P(sim)\$LCCClassesToReplaceNN</code>. The default layer used, if not supplied, is SCANFI-derived data product for 2020 updated to use NTEMS land cover codes. See https://open.canada.ca/data/en/dataset/18e6a919-53fd-41ce-b4e2-44a9707c52dc for SCANFI metadata. The metadata (res, proj, ext, origin) need to match rasterToMatch_biomassParam.</p>	NA

objectName	objectClass	desc	sourceURL
rstWetland	SpatRaster	Site layer on rasterToMatch_biomassParam: non-zero where the ground is wetland . Used to add NTEMS classes 80 (wetland) and 81 (treed wetland) to rstLCC, and as the site axis of P(sim)\$stratumType = 'siteComposition'. If not supplied and P(sim)\$wetlandSource is 'CWIM', built from the Canadian Wetland Inventory Map v3A (bog, fen, marsh and swamp are wet).	NA
rstCanopyHeight	SpatRaster	Canopy height (m) on rasterToMatch_biomassParam. Used only to estimate P(sim)\$deciduousCoverWeight, as one of the two structural controls that let composition be compared between stands carrying the same amount of structure. Defaults to SCANFI's own height layer for P(sim)\$dataYear when P(sim)\$fitDeciduousCoverWeight is TRUE and P(sim)\$dataSource is SCANFI.	NA
rstCanopyClosure	SpatRaster	Canopy closure (percent) on rasterToMatch_biomassParam. The other structural control for P(sim)\$deciduousCoverWeight; see rstCanopyHeight.	NA
rasterToMatch	SpatRaster	A raster of the studyArea in the same resolution and projection as rawBiomassMap. This is the scale used for all outputs for use in the simulation. If not supplied will be forced to match the default rawBiomassMap.	NA

objectName	objectClass	desc	sourceURL
rasterToMatch_biomassParam	SpatRaster	A raster of the studyArea_biomassParam in the same resolution and projection as rawBiomassMap. This is the scale used for all inputs for use in the simulation. If not supplied will be forced to match the default rawBiomassMap.	NA
rawBiomass Map	SpatRaster	total biomass raster layer in study area . Defaults to the Canadian Forestry Service, National Forest Inventory, SCANFI-derived total aboveground biomass map from 2020 (in tonnes/ha), unless dataYear != 2020. See https://open.canada.ca/data/en/dataset/18e6a919-53fd-41ce-b4e2-44a9707c52dc for metadata.	NA
speciesLayers	SpatRaster	cover percentage raster layers by species in Canada species map. Defaults to the Canadian Forestry Service, National Forest Inventory, SCANFI-derived species cover maps from 2020 using a cover threshold of 10 - see https://open.canada.ca/data/en/dataset/18e6a919-53fd-41ce-b4e2-44a9707c52dc for metadata	NA

objectName	objectClass	desc	sourceURL
speciesTable	data.table	a table of invariant species traits with the following trait columns: 'species', 'Area', 'longevity', 'sexualmature', 'shadetolerance', 'firtolerance', 'seeddistance_eff', 'seeddistance_max', 'resproutprob', 'resproutage_min', 'resproutage_max', 'postfireregen', 'leaflongevity', 'wooddecayrate', 'mortalityshape', 'growthcurve', 'leafLignin', 'hardsoft'. Names can differ, but not the column order. Default is from Dominic Cyr and Yan Boulanger's project.	https://raw.githubusercontent.com/dcyr/LANDIS-II_I_A_generalUseFiles/master/speciesTraits.csv
sppColorVect	character	named character vector of hex colour codes corresponding to each species	NA
sppEquiv	data.table	table of species equivalencies. See <code>LandR::sppEquivalencies_CA</code> .	NA
sppNameVector	character	an optional vector of species names to be pulled from <code>sppEquiv</code> . Species names must match <code>P(sim)\$sppEquivCol</code> column in <code>sppEquiv</code> . If not provided, then species will be taken from the entire <code>P(sim)\$sppEquivCol</code> column in <code>sppEquiv</code> . See <code>LandR::sppEquivalencies_CA</code> .	NA

objectName	objectClass	desc	sourceURL
standAgeMap	SpatRaster	stand age map in study area. Must have a 'imputedPixID' attribute (a vector of pixel IDs) indicating which pixels suffered age imputation. If no pixel ages were imputed, please set this attribute to <code>integer(0)</code> . Defaults to the Canadian Forestry Service, National Forest Inventory, SCANFI-derived biomass map from 2020, unless <code>dataYear != 2020</code> . See < https://open.canada.ca/data/en/dataset/18e6a919-53fd-41ce-b4e2-44a9707c52dc > for metadata.	NA
studyArea	sf	sf polygon or terra SpatVector to use as the study area - nrow must be one	NA
studyArea_biomassParam	sf	Polygon to use as the parametrisation study area. Must be provided by the user. Note that <code>studyArea_biomassParam</code> is only used for parameter estimation, and can be larger than the actual study area used for LandR simulations (e.g., larger than <code>studyArea</code> in <code>LandR Biomass_core</code>).	NA

3.2.6 List of parameters

Table 3.3 lists all parameters used in *Biomass_borealDataPrep* and their detailed information. All have default values specified in the module's meta-data.

Of these parameters, the following are particularly important:

Estimation of simulation parameters

- `biomassModel` – the statistical model (as a function call) used to estimate `maxB` and `maxANPP`.
- `coverModel` – the statistical model (as a function call) used to estimate `SEP`.
- `fixModelBiomass` – determines whether `biomassModel` is re-fit when convergence issues arise.
- `imputeBadAgeModel` – model used to impute ages when they are missing, or do not match the input cover and biomass data. Not to be confounded with correcting ages from fire data
- `subsetDataAgeModel` and `subsetDataBiomassModel` – control data subsampling for fitting the `imputeBadAgeModel` and `biomassModel`, respectively
- `exportModels` – controls whether `biomassModel` or `coverModel` (or both) are to be exported in the simulation `simList`, which can be useful to inspect the fitted models and report on statistical fit.
- `sppEquivCol` – character. the column name in the `speciesEquivalency` `data.table` that defines the naming convention to use throughout the simulation.

Data processing

- `forestedLCCClasses` and `LCCClassesToReplaceNN` – define which land-cover classes in `rstLCC` are forested and which should be reclassified to forested classes, respectively.
- `deciduousCoverWeight` and `fitDeciduousCoverWeight` – the first is the adjustment factor for broadleaf species cover to biomass relationships; the second decides whether it is estimated from `studyArea_biomassParam` (the

default) instead, in which case the first is only the fallback for a study area that cannot identify it.

TABLE 3.3: List of *Biomass_borealDataPrep* parameters and their description.

paramName	paramClass	default	min	max	paramDesc	cacheKey
biomassModel	call	lme4::lm...	NA	NA	Model and formula for estimating biomass (B) from ecoregionGroup (currently ecoregionLayer LandCoverClass), speciesCode, logAge (gives a downward curving relationship), and cover. Defaults to a LMEM, which can be slow if dealing with very large datasets (e.g., 36,000 points takes 20 minutes). For faster fitting try <code>P(sim)\$subsetDataBiomassModel == TRUE</code> , or <code>quote(RcppArmadillo::fastLm(formula = B ~ logAge speciesCode ecoregionGroup + cover speciesCode ecoregionGroup))</code> . A custom model call can also be provided, as long as the 'data' argument is NOT included.	TRUE
coverModel	call	glm, cbi....	NA	NA	Model and formula used for estimating cover from ecoregionGroup and speciesCode and potentially others. Defaults to a GLMEM if there are > 1 grouping levels. A custom model call can also be provided, as long as the 'data' argument is NOT included	TRUE
earliestFireYear	integer	1950	NA	NA	if using fires to impute stand age and biomass, the earliest year for which fire data should be obtained	TRUE
fixModelBiomasslogical	FALSE	FALSE	NA	NA	should biomassModel be fixed in the case of non-convergence? Only scaling of variables and attempting to fit with a new optimizer (bobyqa, see ?lme4) are implemented at this time.	TRUE

paramName	paramClass	default	min	max	paramDesc	cacheKey
subsetDataAttempts	integer	3	1	10	How many times should <code>biomassModel</code> be attempted to fit with a new data subset in case of non-convergence? Each time, the data is resampled (if <code>subsetDataBiomassModel</code> = <code>TRUE</code>) and the model re-fit with the original data, scaled variables and/or a different optimizer if <code>fixModelBiomass</code> = <code>TRUE</code> . Model refitting with original data, rescaled variables and/or a new optimizer occurs up to three times for each data subset, regardless of this parameter's value.	TRUE
subsetDataBiomassModel	integer	500	NA	NA	the number of samples per <code>ecoregionGroup</code> x <code>speciesCode</code> to use when subsampling the biomass data model (<code>biomassModel</code>). Can be <code>TRUE</code> / <code>FALSE</code> / <code>NULL</code> or numeric; if <code>TRUE</code> , uses <code>LandR::subsetDataSize()</code> . If <code>FALSE</code> / <code>NULL</code> no subsetting is done. Default: <code>LandR::subsetDataSize()</code> , i.e. option <code>LandR.subsetDataSize(500)</code> .	TRUE

paramName	paramClass	default	min	max	paramDesc	cacheKey
deciduousCoverWeight	Numeric	0.93	NA	NA	How much biomass a unit of deciduous cover carries, relative to a unit of conifer cover: the weight <code>LandR::partitionBiomass()</code> applies to deciduous cover before it splits a pixel's <code>totalBiomass</code> among its cohorts. Broadleaf crowns are wider per unit of wood, so this is expected below 1. This default is only a fallback. With <code>fitDeciduousCoverWeight = TRUE</code> (the default) it is estimated from the study area and this value is not used; it is kept for landscapes that cannot identify it – too little deciduous cover, too little variation in deciduous share between pixels, or no canopy height/closure layers. 0.93 is the mean of the fit on the three landscapes tested so far that can identify it (two Alberta boreal mixedwood areas and NW Ontario boreal shield: 0.90, 1.00, 0.91); it replaces 0.8418911, estimated from NWT data on 2020-03-18 by an estimator since removed.	TRUE
fitDeciduousCoverWeight	Logical	TRUE	NA	NA	If TRUE (default), estimate <code>P(sim)\$deciduousCoverWeight</code> from this study area rather than using the parameter's value. Needs <code>sim\$rstCanopyHeight</code> and <code>sim\$rstCanopyClosure</code> , which default to SCANFI's own layers; without them, or where there is too little deciduous cover to identify it, the parameter value is kept and a message says so. The fit costs a few seconds.	TRUE

paramName	paramClass	default	min	max	paramDesc	cacheKey
adjustAgeAndLongevity	logical	FALSE	NA	NA	Adjust species longevity to the ages observed on the landscape. Cohort ages are capped at 0.9 longevity. Any cohort ages exceeding this threshold are lowered using a smoothed function. If FALSE, no adjustments are applied to age or longevity.	TRUE
dataSource	character	SCANFI	NA	NA	Source for species cover, biomass, age, and landcover data used to initialize cohorts. kNN (2001, 2011) and SCANFI (V2: every 5 years, 1985-2025) provide all necessary layers. Mixing multiple datasets requires additional raster geoprocessing and is not recommended.	TRUE
dataYear	numeric	2020	NA	NA	the year for which dataSource data will be fetched for use with the module. For SCANFI, any year from 1985 to 2025 in steps of 5; LandR::prepRawBiomassMap() stops on a year the source does not provide.	TRUE
ecoregionLayerField	character		NA	NA	the name of the field used to distinguish ecoregions, if supplying a polygon. Defaults to NULL and tries to use 'ECODISTRIC' where available (for legacy reasons), or the row numbers of sim\$ecoregionLayer. If this field is not numeric, it will be coerced to numeric.	TRUE

paramName	paramClass	default	min	max	paramDesc	cacheKey
exportModels	character	none	NA	NA	Controls whether models used to estimate maximum B/ANPP (biomassModel) and species establishment (coverModel) probabilities are exported for posterior analyses or not. This may be important when models fail to converge or hit singularity (but can still be used to make predictions) and the user wants to investigate them further. Can be set to 'none' (no models are exported), 'all' (both are exported), 'biomassModel' or 'coverModel'. BEWARE: because this is intended for posterior model inspection, the models will be exported with data, which may mean very large simList(s)!	TRUE
floorMaxBAtObserved	logical	TRUE	NA	NA	If TRUE, a maxB that the fit clamped to 0 is replaced by the 95th percentile of the biomass the species was observed at in that ecoregionGroup: a negative fit would otherwise stop a species growing where it demonstrably grows. A percentile rather than the maximum, so one freak cohort cannot set the ceiling. Fitted (positive) values are never changed. maxANPP is recomputed for any replaced row (maxB / 30).	TRUE
forestedLCCClasses	numeric	81, 210,....	0	NA	The classes in the rstLCC layer that are 'treed' and will therefore be run in Biomass_core. Defaults to forested classes in NTEMS map (210 conif, 220 deciduous, 230 mixed) plus LandR-generated 240 class, which is recently disturbed forest.	TRUE

144

3 LandR Biomass_borealDataPrep Module

paramName	paramClass	default	min	max	paramDesc	cacheKey
imputeBadAgeModel	lm	lme4::lm..NA	NA	NA	Model and formula used for imputing ages that are either missing or do not match well with biomass or cover. Specifically, if biomass or cover is 0, but age is not, or if age is missing (NA), then age will be imputed. Note that this is independent from replacing ages inside fire perimeters (see <code>P(sim)\$overrideAgeInFires</code>). Defaults to <code>LandR::imputeBadAgeModelDefault()</code> , whose response is <code>log(age)</code> , so an imputed age can never come back negative and be clamped to 0.	TRUE
landis	logical	FALSE	NA	NA	If TRUE, run in 'LANDIS mode': collapse the forested land-cover (<code>rstLCC</code>) classes (<code>P(sim)\$forestedLCCClasses</code>) to a single class so that <code>ecoregionGroup</code> is defined by <code>ecoregion</code> only, not <code>ecoregion x LCC</code> . <code>maxB</code> , <code>maxANPP</code> and species establishment probability are then estimated per <code>ecoregion</code> , as expected by LANDIS-II Biomass Succession. Default FALSE preserves the standard <code>ecoregion x LCC</code> behaviour.	TRUE

paramName	paramClass	default	min	max	paramDesc	cacheKey
LCCClassesToReplaceNN	numeric	240	NA	NA	This will replace these classes on the landscape with the closest forest class <code>P(sim)\$forestedLCCClasses</code> . If the user is using the LCC 2005 land-cover data product for <code>rstLCC</code> , then they may wish to include 36 (cities – if running a historic range of variation project), and 34:35 (burns) Since this is about estimating parameters for growth, it doesn't make any sense to have unique estimates for transient classes in most cases. If no classes are to be replaced, pass <code>'LCCClassesToReplaceNN' = numeric(0)</code> when supplying parameters.	TRUE
LCCClassesToReplaceMethod	character	NA	NA	NA	Passed to <code>LandR::convertUnwantedLCC()</code> as its method argument, controlling how each <code>P(sim)\$LCCClassesToReplaceNN</code> pixel picks among the available classes in its neighbourhood. Both options weight the classes by their local abundance and differ only in reproducibility. 'nearestRandom' (default) draws from the RNG, so replicates differ – but note that <code>Cache()</code> does not key on RNG state, so a cached call replays a single draw unless the seed is part of the cache key. 'nearestWeighted' instead keys the draw on the pixel's ground position, making it deterministic and seed-free, and giving the same answer on a grid-aligned crop of the study area as on the full extent. See <code>?LandR::convertUnwantedLCC</code> .	TRUE
minCoverThreshold	numeric	5	0	100	Pixels with total cover that is equal to or below this number will be omitted from the dataset	TRUE

paramName	paramClass	default	min	max	paramDesc	cacheKey
minSpeciesEcoregionShare	numeric	0.07	0	1	Minimum share of an ecoregion's vegetated pixels in which a species must have cover above <code>minCoverThreshold</code> to be kept in that ecoregion, within <code>studyArea_biomassParam</code> . Below it, the species is removed from every pixel of the ecoregion (its cover goes to the other species), so it also gets <code>establishprob = 0</code> and no <code>maxB/maxANPP</code> there. 0.07 removes western redcedar from the BC mountain hemlock (MH) zone in every ELF (highest: 6.7%) and from ESSF (highest: 2.2%), and keeps it in CWH and in ICH (lowest kept: 7.3%). 0 turns it off.	TRUE
minRelativeBFunction	LandR::mfn	NA	NA	NA	A quoted function that makes the table of min. relative B determining a stand shade level for each <code>ecoregionGroup</code> . Using the internal object <code>pixelCohortData</code> is advisable to access/use the list of <code>ecoregionGroups</code> per pixel. The function must output a <code>data.frame</code> with 6 columns, named <code>ecoregionGroup</code> and 'X1' to 'X5', with one line per <code>ecoregionGroup</code> code, and the min. relative biomass for each stand shade level X1-5. The default function uses values from LANDIS-II available at: https://github.com/dcyr/LANDIS-II_IA_generalUseFiles/blob/master/LandisInputs/BSW/biomass-succession-main-inputs_BSW_Baseline.txt and applies them to all ecolocations (<code>ecoregionGroup</code> codes).	TRUE

paramName	paramClass	default	min	max	paramDesc	cacheKey
initialB	numeric	10	1	NA	Initial biomass values of new age-1 cohorts. Used here as the starting biomass of the young-cohort spin-up; must match Biomass_core's initialB. If NA or NULL, initial biomass will be calculated as in LANDIS-II Biomass Suc. Extension (see Scheller and Miranda, 2015 or ?LandR::initiateNewCohorts)	TRUE
omitNonTreedException	logical	TRUE	FALSE	TRUE	Should this module use only treed pixels, as identified by P(sim)\$forestedLCCClasses?	TRUE
overrideAgeInFires	logical	TRUE	NA	NA	should stand age values inside fire perimeters be replaced with number of years since last fire?	TRUE
overrideBiomassInFires	logical	TRUE	NA	NA	should B values be re-estimated using Biomass_core for pixels within the fire perimeters for which age was replaced with time since last fire? Ignored if P(sim)\$overrideAgeInFires = FALSE. See firePerimeters input object and P(sim)\$overrideAgeInFires for further detail.	TRUE
pixelGroupAgeClass	numeric		NA	NA	When assigning pixelGroup membership, this defines the resolution of ages that will be considered 'the same pixelGroup', e.g., if it is 10, then 6 and 14 will be the same	TRUE
pixelGroupBiomassClass	numeric	100	NA	NA	When assigning pixelGroup membership, this defines the resolution of biomass that will be considered 'the same pixelGroup', e.g., if it is 100, then 5160 and 5240 will be the same	TRUE
rmImputedPixels	logical	FALSE	NA	NA	Should sim\$imputedPixID be removed from the simulation?	TRUE

148

3 LandR Biomass_borealDataPrep Module

paramName	paramClass	default	min	max	paramDesc	cacheKey
speciesUpdateFunction	list	LandR::s...	NA	NA	Unnamed list of (one or more) quoted functions that updates species table to customize values. By default, LandR::speciesTableUpdate is used to change longevity and shade tolerance values, using values appropriate to Boreal Shield West (BSW), Boreal Plains (BP) and Montane Cordillera (MC) ecoprovinces (see ?LandR::speciesTableUpdate for details). Set to NULL if default trait values from speciesTable are to be kept instead. The user can supply other or additional functions to change trait values (see LandR::updateSpeciesTable)	TRUE
sppEquivCol	character	LandR	NA	NA	The column in sim\$speciesEquivalency data.table to use as a naming convention.	TRUE
speciesTableArea	character	BSW, BP, MC	NA	NA	One or more of the Ecoprovince short forms that are in the speciesTable file, e.g., BSW, MC etc. Default is good for Alberta and other places in the western Canadian boreal forests.	TRUE
stratumMinPixels	numeric	100	0	NA	Under stratumType = 'siteComposition', the minimum number of estimation pixels an ecoregion x site x composition stratum needs to keep its own parameters. A thinner stratum loses its composition first (pooled codes 290 upland, 890 wet) and, if that pool is still thin, its site too (990). 0 pools nothing.	TRUE

paramName	paramClass	default	min	max	paramDesc	cacheKey
stratumType	character	landcover	NA	NA	How pixels are grouped (ecoregionGroup) for estimating maxB, maxANPP and establishment probability. 'landcover': ecoregion x land-cover class, with the NTEMS wetland classes 80/81 added from rstWetland – one axis, as NTEMS has it. 'siteComposition': ecoregion x site x composition, so a species on upland and on wet ground get separate parameters. Codes: upland 210/220/230, wet 810/820/830; pooled 290/890/990 (see stratumMinPixels); class-240 pixels with no species cover 240/840.	TRUE
subsetDataAgeModel	numeric	500	NA	NA	the number of samples per group to use when subsampling the age data model. Can be TRUE/FALSE/NULL or numeric; if TRUE, uses LandR::subsetDataSize(). If FALSE/NULL no subsetting is done. Default: LandR::subsetDataSize(), i.e. option LandR.subsetDataSize(500).	TRUE
successionTimeStep	numeric	10	NA	NA	defines the simulation time step, default is 10 years	TRUE
useCloudCacheForStats	logical	TRUE	NA	NA	Some of the statistical models take long (at least 30 minutes, likely longer). If this is TRUE, then it will try to get previous cached runs from googledrive.	TRUE
vegLeadingProportion	numeric	0.75	0	1	a number that defines whether a species is leading for a given pixel. Default: LandR::leadingSpeciesProp(), i.e. option LandR.leadingSpeciesProp, which takes LandR.mixedwoodProp (0.75) unless set. Setting it in one place moves every module and LandR function together.	TRUE

159

3 LandR Biomass_borealDataPrep Module

paramName	paramClass	default	min	max	paramDesc	cacheKey
wetlandSource	character	CWIM	NA	NA	Where <code>rstWetland</code> comes from when it is not supplied. 'CWIM': the Canadian Wetland Inventory Map v3A via <code>LandR::prepInputs_CWIM()</code> – SCANFI land cover has no wetland classes, so without it treed wetland cannot be told from upland forest. 'none': no site layer; every pixel is treated as upland.	TRUE
.plotInitialTime	numeric	0	NA	NA	This is here for backwards compatibility. Please use <code>.plots</code>	TRUE
.plots	character	NA	NA	NA	This describes the type of 'plotting' to do. See <code>?Plots</code> for possible types. To omit, set to NA	TRUE
.plotInterval	numeric	NA	NA	NA	This describes the simulation time interval between plot events	TRUE
.saveInitialTime	numeric	NA	NA	NA	This describes the simulation time at which the first save event should occur	TRUE
.saveInterval	numeric	NA	NA	NA	This describes the simulation time interval between save events	TRUE
.seed	list		NA	NA	Named list of seeds to use for each event (names). E.g., <code>list('init' = 123)</code> will set <code>seed(123)</code> at the start of the init event and unset it at the end. Defaults to NULL, meaning that no seeds will be set	TRUE
.sslVerify	integer	64	NA	NA	Passed to <code>httr::config(ssl_verifypeer = P(sim)\$sslVerify)</code> when downloading NFI datasets. Set to OL if necessary to bypass checking the SSL certificate (this may be necessary when NFI's website SSL certificate is not correctly configured).	TRUE
.studyAreaName	character	NA	NA	NA	Human-readable name for the study area used. If NA, a hash of <code>studyArea</code> will be used.	TRUE

paramName	paramClass	default	min	max	paramDesc	cacheKey
.useCache	character	inputOb.	NA	NA	Internal. Can be names of events or the whole module name; these will be cached by SpaDES	TRUE
.useCloud	logical	FALSE	NA	NA	should a cloud cache be used for heavy operations	TRUE
.useCacheArgs	list	list(use....	NA	NA	should this event be cloud cached	TRUE

3.2.7 List of outputs

The module produces the following outputs (Table 3.4), which are key inputs of *Biomass_core*.

Tables

- `cohortData` – initial community table, containing corrected biomass (g/m²), age and species cover data, as well as ecolocation and `pixelGroup` information. This table defines the initial community composition and structure used by *Biomass_core*.
- `species` – table of invariant species traits. Will contain the same traits as in `speciesTable` above, but adjusted where necessary.
- `speciesEcoregion` – table of spatially-varying species traits (`maxB`, `maxANPP`, `SEP`).
- `minRelativeB` – minimum relative biomass thresholds that determine a shade level in each pixel. XO-5 represent site shade classes from no-shade (0) to maximum shade (5).
- `sufficientLight` – probability of germination for species shade tolerance (in `species`) and shade level (defined by `minRelativeB`)

Spatial layers

- `biomassMap` – map of initial stand biomass values after adjustments for data mismatches.
- `pixelGroupMap` – a map containing `pixelGroup` IDs per pixel. This defines the initial map used for hashing within *Biomass_core*, in conjunction with `cohortData`.
- `ecoregionMap` – map of ecolocations.

TABLE 3.4: List of *Biomass_borealDataPrep* output objects and their description.

objectName	objectClass	desc
biomassMap	SpatRaster	total biomass raster layer in study area, filtered for pixels covered by cohortData. Units in g/m^2
cohortData	data.table	initial community table, containing corrected biomass (g/m^2), age and species cover data, as well as ecolocation and pixelGroup information. This table defines the initial community composition and structure used by Biomass_core
ecoregion	data.table	ecoregionGroup look up table
ecoregionMap	SpatRaster	ecoregionGroup map that has mapcodes match ecoregion table and speciesEcoregion table
firePerimeters	SpatRaster	As the input object firePerimeters, but potentially cropped/masked/projected to match rasterToMatch_biomassParam
imputedPixID	Integer	A vector of pixel IDs - matching rasterMatch IDs - that suffered data imputation. Data imputation may be in age (to match last fire event post 1950s, or 0 cover), biomass (to match fire-related imputed ages, correct for missing values or for 0 age/cover), land cover (to convert non-forested classes into to nearest forested class)
pixelGroupMap	SpatRaster	initial community map that has mapcodes (pixelGroup IDs) match cohortData
pixelFateDT	data.table	A small table that keeps track of the pixel removals and cause. This may help diagnose issues related to understanding the creation of cohortData.
minRelativeB	data.frame	minimum relative biomass thresholds that determine a shade level in each pixel. x0-5 represent site shade classes from no-shade (0) to maximum shade (5).
modelCover	data.frame	if P(sim)\$exportModels is 'all', or 'cover', fitted cover model, as defined by P(sim)\$coverModel.
modelBiomass	data.frame	if P(sim)\$exportModels is 'all', or 'biomass', fitted biomass model, as defined by P(sim)\$biomassModel

objectName	objectClass	desc
rstLCC	SpatRaster	As the input object <code>rstLCC</code> , but potentially cropped/projected/masked to match <code>rasterToMatch_biomassParam</code>
species	data.table	Table that of invariant species traits. Will have the same traits as the input <code>speciesTable</code> , with values adjusted where necessary.
speciesLayers	SpatRaster	cover percentage raster layers by species in Canada species map. Defaults to the Canadian Forestry Service, National Forest Inventory, SCANFI-derived species cover maps from 2020 using a cover threshold of 10 - see https://open.canada.ca/data/en/dataset/18e6a919-53fd-41ce-b4e2-44a9707c52dc for metadata.
speciesEcoregion	data.table	table of spatially-varying species traits (<code>maxB</code> , <code>maxANPP</code> , <code>establishprob</code>), defined by species and <code>ecoregionGroup</code> (i.e. ecolocation)
standAgeMap	SpatRaster	As the input object <code>standAgeMap</code> , but potentially cropped, projected, masked to match <code>rasterToMatch_biomassParam</code> .
sufficientLight	data.frame	Probability of germination for species shade tolerance (in <code>species</code>) and shade level (defined by <code>minRelativeB'</code>) combinations. Table values follow LANDIS-II test traits available at: https://raw.githubusercontent.com/LANDIS-II-Foundation/Extensions-Succession/master/biomass-succession-archive/trunk/tests/v6.0-2.0/biomass-succession_test.txt

3.2.8 Simulation flow and module events

Biomass_borealDataPrep initialises itself and prepares all inputs, provided it has internet access to retrieve the raw datasets, for parametrisation and use by *Biomass_core*.

The module runs only for one time step and contains The general flow of *Biomass_borealDataPrep* processes is:

1. Preparation of all necessary data and input objects that do not require parameter fitting (e.g., invariant species traits table, creating ecolocations);
2. Fixing mismatched between raw cover, biomass and age data;
3. Imputing age values in pixels where mismatches exist or age data is missing;
4. Construction of an initial `data.table` of cohort biomass and age per pixel (with ecolocation information);
5. Sub-setting pixels in forested land-cover classes and (optional) converting transient land-cover classes to forested classes;
6. Fitting `coverModel`;
7. Fitting `biomassModel` (and re-fitting if necessary – optional);
8. Estimating `maxB`, `maxANPP` and `SEP` per species and ecolocation.
9. (OPTIONAL) Correcting ages in pixels inside fire perimeters and reassigning biomass.

[steps 1-9 are part of the `init` event. Before step 1, the data is downloaded when during the run of the `.inputObjects` function]

10. (OPTIONAL) Plots of `maxB`, `maxANPP` and `SEP` maps (`plot` event);
11. (OPTIONAL) Save outputs (`save` event)

3.3 Usage example

This module can be run stand-alone, but it won't do much more than prepare inputs for `Biomass_core`. Hence, we provide a usage example of this module and a few others in this repository¹⁸ and in Barros et al. [1].

3.4 References

- [1] Ceres Barros et al. “Empowering ecological modellers with a PERFICT workflow: Seamlessly linking data, parameterisation, prediction, validation and visualisation”. In: *Methods in Ecology and Evolution* 14 (2023), pp. 173–188. DOI: 10.1111/2041-210X.14034¹⁹.
- [3] A Beaudoin et al. “Mapping attributes of Canada’s forests at moderate resolution through kNN and MODIS imagery”. In: *Canadian Journal of Forest Research* 44 (2014), pp. 521–532. DOI: 10.1139/cjfr-2013-0401²⁰.
- [4] A Beaudoin et al. *Species composition, forest properties and land cover types across Canada’s forests at 250m resolution for 2001 and 2011*. 2017. DOI: 10.23687/EC9E2659-1C29-4DDB-87A2-6ACED147A990²¹. URL: <http://open.canada.ca/data/en/dataset/ec9e2659-1c29-4ddb-87a2-6aced147a990>.
- [6] P J Burton and S G Cumming. “Potential effects of climatic change on some western Canadian forests, based on phenological enhancements to a patch model of forest succession”. In: *Water, Air and Soil Pollution* 82 (1995), pp. 401–414.
- [11] L Guindon et al. *SCANFI: the Spatialized Canadian National Forest Inventory data product*. 2023. DOI: 10.23687/18E6A919-53FD-41CE-B4E2-44A9707C52DC²². URL: <https://open.canada.ca/data/en/dataset/18e6a919-53fd-41ce-b4e2-44a9707c52dc> (visited on 09/26/2025).

¹⁸https://github.com/CeresBarros/LandRBiomass_publication

¹⁹<https://doi.org/10.1111/2041-210X.14034>

²⁰<https://doi.org/10.1139/cjfr-2013-0401>

²¹<https://doi.org/10.23687/EC9E2659-1C29-4DDB-87A2-6ACED147A990>

²²<https://doi.org/10.23687/18E6A919-53FD-41CE-B4E2-44A9707C52DC>

- [12] Luc Guindon et al. “A new approach for spatializing the Canadian National Forest Inventory (SCANFI) using Landsat dense time series”. en. In: *Canadian Journal of Forest Research* 54.7 (July 2024), pp. 793–815. ISSN: 0045-5067, 1208-6037. DOI: 10.1139/cjfr-2023-0118²³. URL: <https://cdnsiencepub.com/doi/10.1139/cjfr-2023-0118> (visited on 09/26/2025).
- [14] Txomin Hermosilla et al. “Land cover classification in an era of big and open data: Optimizing localized implementation and training data selection to improve mapping outcomes”. en. In: *Remote Sensing of Environment* 268 (Jan. 2022), p. 112780. ISSN: 00344257. DOI: 10.1016/j.rse.2021.112780²⁴. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0034425721005009> (visited on 10/02/2025).
- [15] E. H. (Ted) Hogg, J. P. Brandt, and M. Michaelian. “Impacts of a regional drought on the productivity, dieback, and biomass of western Canadian aspen forests”. In: *Canadian Journal of Forest Research* 38.6 (June 2008), pp. 1373–1384. ISSN: 0045-5067, 1208-6037. DOI: 10.1139/X08-001²⁵. URL: <http://www.nrcresearchpress.com/doi/10.1139/X08-001>.
- [19] Christian Messier, Sylvain Parent, and Yves Bergeron. “Effects of overstory and understory vegetation on the understory light environment in mixed boreal forests”. In: *Journal of Vegetation Science* 9.4 (Aug. 1998), pp. 511–520. ISSN: 11009233. DOI: 10.2307/3237266²⁶. URL: <http://doi.wiley.com/10.2307/3237266>.
- [20] NRCan. *National Ecological Framework for Canada - Terrestrial Ecozones*. 2013.
- [21] Changhui Peng et al. “A drought-induced pervasive increase in tree mortality across Canada’s boreal forests”. In: *Nature Climate Change* 1.9 (Dec. 2011), pp. 467–471. ISSN: 1758-678X, 1758-6798. DOI: 10.1038/nclimate1293²⁷. URL: <http://www.nature.com/articles/nclimate1293>.
- [26] Robert M. Scheller and Brian R. Miranda. *LANDIS-II Biomass Succession v3.2 Extension – User Guide*. 2015.

²³<https://doi.org/10.1139/cjfr-2023-0118>

²⁴<https://doi.org/10.1016/j.rse.2021.112780>

²⁵<https://doi.org/10.1139/X08-001>

²⁶<https://doi.org/10.2307/3237266>

²⁷<https://doi.org/10.1038/nclimate1293>

- [27] Robert M. Scheller and David J. Mladenoff. “A forest growth and biomass module for a landscape simulation model, LANDIS: design, validation, and application”. In: *Ecological Modelling* 180.1 (Dec. 2004), pp. 211–229. DOI: 10.1016/j.ecolmodel.2004.01.022²⁸. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0304380004003837>.
- [32] Mark D. Wilkinson et al. “The FAIR Guiding Principles for scientific data management and stewardship”. In: *Scientific Data* 3.1 (Dec. 2016), p. 160018. ISSN: 2052-4463. DOI: 10.1038/sdata.2016.18²⁹. URL: <http://www.nature.com/articles/sdata201618>.

²⁸<https://doi.org/10.1016/j.ecolmodel.2004.01.022>

²⁹<https://doi.org/10.1038/sdata.2016.18>



4

LandR Biomass_speciesParameters Module

Biomass speciesParameters v. 3.1.0.9000 ¹

Get help Report issues ²

4.0.0.1 Authors:

Ian Eddy ian.eddy@nrcan-rncan.gc.ca³ [aut, cre], Eliot McIntire eliot.mcintire@nrcan-rncan.gc.ca⁴ [aut], Ceres Barros ceres.barros@ubc.ca⁵ [ctb]

This documentation is work in progress. Potential discrepancies and omissions may exist for the time being. If you find any, contact us using the “Get help” link above.

4.1 Module Overview

4.1.1 Quick links

- General functioning
- List of input objects
- List of parameters
- List of outputs

¹https://github.com/PredictiveEcology/Biomass_speciesParameters6c35e5e6ebbadbc673da5ada1b2d1c16ca25afed

²https://github.com/PredictiveEcology/Biomass_speciesParameters/issues

³<mailto:ian.eddy@nrcan-rncan.gc.ca>

⁴<mailto:eliot.mcintire@nrcan-rncan.gc.ca>

⁵<mailto:ceres.barros@ubc.ca>

- Simulation flow and module events

4.1.2 Summary

LandR *Biomass_speciesParameters* (hereafter *Biomass_speciesParameters*) calibrates species growth and mortality trait values used in *Biomass_core*, by matching theoretical species' growth curves obtained with different trait values (see *Simulated species data*) against observed growth curves derived from Permanent Sample Plots (PSP data) across Canada (see *Permanent sample plot data*), to find the combination of trait values that allows a better match to the observed curves. In particular, it calibrates the *growthcurve*, *mortalityshape*, maximum biomass (*maxB*) and maximum aboveground net primary productivity (*maxANPP*) traits (see *Parameter estimation/calibration*), the latter two in conjunction with the module *Biomass_borealDataPrep*.

This module **will not** obtain other traits or parameters used in *Biomass_core* and so it is meant to be used in conjunction with another data/calibration module that does so (e.g., *Biomass_borealDataPrep*). However it can be used stand-alone in an initial developmental phase for easier inspection of the statistical calibration procedure employed.

As of October 29, 2026, the *raw* PSP data used in this module is not freely available, and data sharing agreements must be obtained from the governments of SK, AB, and BC to obtain it. However, the *processed and anonymized* PSP data is provided via a Google Drive folder accessed automatically by the module.

*Google Account is therefore necessary to access the data used for calibration.**

4.1.3 Links to other modules

Biomass_speciesParameters is intended to be used with another data module, like *Biomass_borealDataPrep*, that prepares all other traits and parameters (including *maxB* and *maxANPP*) for *Biomass_core*. See here⁶ for all available modules in the LandR ecosystem and select *Biomass_speciesParameters* from the drop-down menu to see potential linkages.

⁶https://rpubs.com/PredictiveEcology/LandR_Module_Ecosystem

- *Biomass_borealDataPrep*⁷: prepares all parameters and inputs (including initial landscape conditions) that *Biomass_core* needs to run a realistic simulation. Default values/inputs produced are relevant for boreal forests of Western Canada. Used upstream from *Biomass_speciesParameters*;
- *Biomass_core*⁸: core forest dynamics simulation module. Used downstream from *Biomass_speciesParameters*;
- *Biomass_speciesFactorial*⁹: a module that generates theoretical species curves by running thousands of *Biomass_core* simulations on landscapes populated by one or more species, each simulation using a different set of species trait values.

4.2 Module manual

4.2.1 General functioning

Tree cohort growth and mortality in *Biomass_core* are essentially determined by five parameters: *growthcurve*, *mortalityshape*, maximum biomass (*maxB*), maximum aboveground net primary productivity (*maxANPP*) and *longevity*.

The *growthcurve* and *mortalityshape* parameters (called ‘growth curve’ and ‘mortality shape’ in LANDIS-II Biomass Succession Extension v3.2, the base model for *Biomass_core*) strongly modulate the shape of species growth curves and so it is important that they are calibrated to the study area in question.

Also, the growth and mortality equations used in *Biomass_core* are non-linear and their resulting actual biomass accumulation curve is an emergent phenomenon due to competition effects. This means that the ideal trait/parameter values should not be estimated on pure single species growth conditions, as their resulting dynamics will be different in a multi-species context.

Biomass_speciesParameters attempts to address these issues (at least partially) using a “curve-matching” approach. It compares the best fit (ac-

⁷https://github.com/PredictiveEcology/Biomass_borealDataPrep

⁸https://github.com/PredictiveEcology/Biomass_core

⁹https://github.com/PredictiveEcology/Biomass_core

cording to their AIC) of three non-linear forms (Chapman-Richard's, Gompertz, and a logistic form) fitted to permanent sample plot (PSP) data to a large collection of theoretical (i.e. simulated) species curves, each representing a different set of the five key parameters that govern biomass increment in `Biomass_core: growthcurve`, `mortalityshape`, the ratio of `maxANPP` to `maxB`, and `longevity`. This library of curves is produced by the *Biomass_speciesFactorial* module.

Biomass_speciesParameters generally follows other LandR data modules, like *Biomass_boreaDataPrep*, which also attempts to calibrate previously estimated spatially varying species traits such as `maxB` and `maxANPP` from the input data layers.

4.2.1.1 Permanent sample plot data

Biomass_speciesParameters can use all the PSP data available (note that it may span several thousands of kilometres), or select the data based on a shapefile (`studyAreaANPP`; see [List of input objects](#)).

By default, all available PSP is obtained, via the `PSPdataTypes` parameter. The particular data sets will depend on the version of `ianmseddy/PSPclean` that is installed (assuming requisite file-sharing permissions are available). This may include data from the provinces of BC, AB, SK, ON, QC, and NB, as well as the National Forest Inventory. If no file-sharing agreements are in place, the dummy option can be used. This will obtain data from BC, AB, SK, and the NFI that were previously treated for errors and standardized into a single data set with the exact location and identifying attributes anonymized. However, it should be noted that data for BC, Quebec, New Brunswick, and the NFI are freely available, thus no file-sharing agreement is necessary for these jurisdictions.

The data include individual species, diameter at breast height (DBH), and sometimes tree height measurements for each tree in a plot, as well as stand age. As part of the standardization process, dead trees were removed from the data set. Tree biomass was then per estimated species using either a DBH-only model or a DBH-height model from Lambert, Ung, and Raulier [16], in g/m^2 .

Note that the model used to calculate biomass can also be changed to Ung, Bernier, and Guo [30] via the `P(sim)$biomassModel` module parameter (see [list of parameters](#)).

4.2.1.2 Simulated species data

The *Biomass_speciesFactorial* module was used to create a library of theoretical species curves (biomass accumulation curves, to be more precise) to which the best non-linear model form fit to the PSP-biomass will be matched for each species and species combinations in the study area landscape. The library of curves was created by running several *Biomass_core* simulations with no reproduction, competition, disturbance, or dispersal effects. The species in the simulations encompassed the full factorial of traits, growing alone and in competition with one cohort of each other combination. Each simulation differed in the combination of species trait values that influence growth and mortality dynamics, namely: growthcurve, mortalityshape, longevity, maxANPP and maximum biomass (maxBiomass, not to be confused with the data-driven maxB which is later calibrated).

The values for maxANPP were explored via the mANPPproportion, the ratio of maxANPP to maxBiomass (the parameter used for theoretical curves), as it reflects their relationship. While any factorial combination is possible to run using the module *Biomass_speciesFactorial*, the default objects used by *Biomass_speciesParameters* utilized growthcurve values ranging from 0.65 to 0.85 in increments of 0.02, mortalityshape values of 20, 22, and 24, mANPPproportion values from 3.5 to 6.5 in increments of 0.5, and longevity values of 125 to 700 in increments of 25, all of which total over 6.5 million unique combinations of traits.

Results from these simulations were compiled into a table (cohortDataFactorial ; see [List of input objects](#)) that is accessed by *Biomass_speciesParameters*, so that the module can be run without needing to re-simulate the theoretical curves.

4.2.1.3 Parameter estimation/calibration

Biomass_speciesParameters calibrates growthcurve, mortalityshape and mANPPproportion by matching the theoretical species curves produced by *Biomass_speciesFactorial* (cohortDataFactorial) against observed species growth curves from the PSP data.

The parameter P(sim)\$speciesFittingApproach determines which of four possible fitting approaches to use: all, single, pair-wise, or focal. The all approach combines all PSPs into a single species, thus parameterizing all species identically. This is not intended to have an ecological application. The "single" approach is the earliest method, where each species utilizes

plots where 50% or more of the biomass is composed solely of that species. Non-linear models for each species are fit to these observations independently of each other, with each species possessing its own subset of plots. However, this approach is unable to accurately characterize the competitive effects that arise when multiple species occupy the same plot. Therefore it is suitable only for characterizing species that grow in pure stands, and has been retained largely for backwards compatibility.

The "pairwise" approach retains all plots where exactly two species each represent more than 20% of the total plot biomass, e.g. a plot with 40% *Pinus contorta* and 39% *Populus tremuloides* is retained only if the remaining 21% of biomass is composed of more than one species. Then, for each combination of 2 species-of-interest, separate non-linear models are fit for each species. #TODO: I believe the models are ultimately combined but this might not happen until traits are selected. Of the three approaches, the pairwise is best able to account for competition, but is most sensitive to data availability and can be confounded by combinations of species that seldom occur together.

The third approach `speciesFittingApproach`, "focal", is the default. This approach is similar to "pairwise", but combines the species that are *not* the species-of-interest (or 'focal' species). For example, in a plot with biomass composition of 20% *Pinus contorta*, 40% *Populus tremuloides*, and 21% *Picea mariana*, when fitting the *P. contorta* equations, the biomass of *P. tremuloides* and *P. mariana* are combined, and when fitting *P. tremuloides*, the biomass of *P. contorta* and *P. mariana* are combined. In comparison to the pairwise approach, this approach sacrifices some detail but is less sensitive to data quality and availability.

Before calculating the *observed* species growth curves (i.e., the best of three non-linear forms to match PSP data), the module subsets the PSP data to stand ages below the 95th percent quantile for all species (this can be changed via the `P(sim)$quantileAgeSubset` module parameter), as records for larger age classes were limited and constituted statistical outliers. In some species, changing the quantile value may improve results, however. Two examples are *Pinus banksiana* and *Populus sp* (in western Canada), for which using the 99th percent quantile improved the models, because these are short-lived species for which data at advanced ages is scarce. In addition, weights are added at the origin (age = 0 and biomass = 0) to force the intercept to be essentially at 0 age and 0 biomass.

The best fit of three non-linear forms, for each focal species, is then calcu-

lated. Focal species are defined as either 50% of dominance in the plot, or 20% if we are looking to capture the multi-species dynamics (currently the default). Three growth model forms are then fit to the observations for the focal species: a Chapman-Richard's form [Equation (4.1); see, e.g., Coble and Lee [8]], a Gompertz form (Equation (4.2)) and a Logistic form [Equation (4.3); see Fekedulegn, Mac Siurtain, and Colbert [10] for a complete overview of these equations]. Multiple tries using the estimation methods from the `robustbase::nlrob` function for each form are used, and the best model fit is selected via Akaike Information Criterion (AIC).

$$B \sim A \times (1 - e^{-k \times age})^p \quad (4.1)$$

$$B \sim A \times e^{-k \times e^{-p \times age}} \quad (4.2)$$

$$B \sim \frac{A}{1 + k \times e^{-p \times age}} \quad (4.3)$$

Species biomass (B) is estimated as a function of stand age (age), with the best values of the A , k and p parameters to fit the PSP data.

It is possible that some selected species do not have enough data to allow for model convergence. In this case, `Biomass_speciesParameters` skips parameter calibration. Consequently the module will interpolate the `maxANPPproportion`, `growthcurve`, and `mortalityshape` from the respective means of other species. These species are tracked via the `source` column in `sim$species`, which will be one of "interpolated" or "estimated". This mechanism ensures these species remain competitive, as the parameterized traits are often significant departures from defaults used in many LANDIS-II applications to Canada's boreal forests.

After each species best fit is selected (using AIC), `Biomass_speciesParameters` compares it to the library of theoretical curves, and picks the best one based on maximum likelihood. This best theoretical curve will be associated with a given combination of `growthcurve`, `mortalityshape` and `maxANPPproportion` values, which are then used directly as the calibrated values, in case of `growthcurve` and `mortalityshape`, or to calibrate `maxANPP` in the case of `maxANPPproportion` (see below).

Because simulated growth curves never achieve the maximum biomass parameter (the `maxBiomass` parameter set to 5000 for all simulations of theoretical species curves, or the `maxB` parameter in `Biomass_core` simulations), it

acts as an asymptotic limit that reflects the potential maximum biomass for a species in an ecolocation (ecological zone and land cover combination).

Biomass_speciesParameters uses the ratio between the potential maximum biomass (`maxBiomass`, always 5000) to the achieved maximum biomass in the theoretical curves, to rescale `maxB`. This ratio is called the `inflationFactor` and it is multiplied by `maxB` values previously estimated from data (e.g. by *Biomass_borealDataPrep*). This way, species simulated in *Biomass_core* are able to achieve the maximum observed biomass used to *initially* estimate `maxB`.

Finally, the module calibrates `maxANPP` using the `manPPproportion` value from the best matching theoretical growth curve as:

$$\text{maxB} \times \frac{\text{manPPproportion}}{100} \quad (4.4)$$

where `maxB` is the already (re-)calibrated version.

4.2.2 List of input objects

The full list of input objects required by the module is presented below (Table 4.1). The input `studyAreaANPP` (the study area used extract the PSP data from) is optional, and therefore no default is supplied. All other input objects have internal defaults, but the user may need to request access to their online files.

Of these inputs, the following are particularly important and deserve special attention:

Spatial layers

- ``studyAreaANPP`` -- shapefile. An ``sf`` or ``SpatVector`` object determining the geographic extent of the PSP data.

• Tables

- `speciesTableFactorial` and `cohortDataFactorial` – a tables of species trait combinations and the theoretical species growth curve data (respectively)
- `PSPmeasure_sppParams`, `PSPplot_sppParams` and `PSPgis_sppParams` – tree measurement, biomass growth and geographical data of the PSP datasets used to build observed species growth curves.

- `species` – a table of invariant species traits that may have been produced by another module. It **must** contain the columns ‘species’, ‘growthcurve’ and ‘mortality shape’, whose values will be calibrated.
- `speciesEcoregion` – table of spatially-varying species traits that may have been produced by another module. It **must** contain the columns ‘speciesCode’, ‘maxB’ and ‘maxANPP’ and ‘ecoregionGroup’ (the ecolocation ID). ‘maxB’ and ‘maxANPP’ values will be calibrated by species.

TABLE 4.1: List of *Biomass_speciesParameters* input objects and their description.

objectName	objectClass	desc	sourceURL
cohortData Factorial_path	fs_path	Path where the cohortDataFactorial object is saved as an arrow dataset. A large cohortData table (sensu Biomass_core) with columns age, B, and speciesCode that joins with speciesTableFactorial. See PredictiveEcology/Biomass_factorial for further information.	https://drive.google.com/file/d/1NH7OpAnWtLyO8JVnhwdMJakOyapBnuBH/view?usp=sharing
PSPmeasure _sppParams	data.table	Merged PSP and TSP individual tree measurements. Must include the following columns: MeasureID, OrigPlotID1, MeasureYear, TreeNumber, Species, DBH and PSP, where Species corresponds to species names in LandR::sppEquivalencies_CA\$Latin_full. Defaults to randomized PSP data stripped of real plotIDs	https://drive.google.com/file/d/1LmOaEtCZ6EBellAm6ttfLqBqQnQu4Ca7/view?usp=sharing

objectName	objectClass	desc	sourceURL
PSPplot_sppParams	data.table	Merged PSP and TSP plot data. Defaults to randomized PSP data stripped of real plotIDs. Must contain columns MeasureID, MeasureYear, OrigPlotID1, and baseSA, the latter being stand age at year of first measurement	https://drive.google.com/file/d/1LmOaEtCZ6EBellAm6ttfLqBqQnQu4Ca7/view?usp=sharing
BECzonesBC	sf	Optional. BC BEC zone polygons, with a ZONE column, used to exclude the zones of excludeBECzonesHybridSpruce from the hybrid-spruce relabel of PSPs. If not supplied, they are fetched from the BC Data Catalogue (WHSE_FOREST_VEGETATION.BEC_BIOGEOCLIMATIC_POLY); if that fails there is no relabel.	NA

objectName	objectClass	desc	sourceURL
PSPgis_spp Params	sf	Plot location sf object. Defaults to P SP data stripped of real plotIDs/location. Must include field OrigPlotID1 for joining to PSPplot object	https://drive.google.com/file/d/1LmOaEtCZ6EBellAm6ttfLqBqQnQu4Ca7/view?usp=sharing
species	data.table	A table of invariant species traits with the following trait columns: 'species', 'Area', 'longevity', 'sexualmature', 'shade tolerance', 'firetolerance', 'seeddistance_eff', 'seeddistance_max', 'resproutprob', 'mortalityshape', 'growthcurve', 'resproutage_min', 'resproutage_max', 'postfireregen', 'wooddecayrate', 'leaflongevity', 'leaflignin', and 'hardsoft'. Only 'growthcurve', 'hardsoft', and 'mortalityshape' are used in this module. Default is from Dominic Cyr and Yan Boulanger's applications of LANDIS-II	https://raw.githubusercontent.com/dcyr/LANDIS-II_I_A_generalUseFiles/master/speciesTraits.csv
speciesEco region	data.table	Table of spatially-varying species traits (maxB, maxANPP, establishprob), defined by species and ecoregionGroup). Defaults to a dummy table based on dummy data of biomass, age, ecoregion and land cover class.	NA

objectName	objectClass	desc	sourceURL
speciesTableFactorial_path	fs_path	Path where the speciesTableFactorial object is saved as an arrow dataset. A large species table (sensu Biomass_core) with all columns used by Biomass_core, e.g., longevity, growthcurve, mortalityshape, etc., when it was used to generate cohortDataFactorial. See PredictiveEcology/Biomass_factorial for further information.	https://drive.google.com/file/d/1NH7OpAnWtLyO8JVnhwdMJakOyapBnuBH/
sppEquiv	data.table	Table of species equivalencies - see ?LandR::sppEquivalencies_CA. Traits will be estimated for each unique entry in the sppEquivCol column.	NA
sppEquivLong	data.table	The full table of species equivalencies - see ?LandR::sppEquivalencies_CA. Biomass will be estimated for each species based on the sp_Biomass_eq' column, which uses pemisc::biomassCalculation' to derive AGB from DBH and height (based on the equations from https://doi.org/10.1139/x05-112). The full table is used to improve stand biomass estimates even if some species are not of interest.	NA

studyAreaA NPP	sf	Optional study area used to crop PSP data before building growth curves. If supplied, an ecoregion-scale object is recommended, at a minimum.	NA	174
----------------	----	---	----	-----

4.2.3 List of parameters

The full list of parameters used by the module is presented below (Table 4.2), all of which have default values specified in the module's metadata.

Of these parameters, the following are particularly important:

Calibration parameters

- ``standAgesForFitting`` -- determines the range of ages for which the fit of the growth curve is linear model. It should be a subset of the full growth curve, because there is extremely limited data before age 20, as well as towards the end of a species' longevity.
- ``speciesFittingApproach`` -- should the calibration take into account species growing in single- or multi-species context?
- ``quantileAgeSubset`` -- upper quantile age value used to subset PSP data. It can be a vector named by species, or a single numeric. It is used to limit the outsized influence of the distribution. However, it will have varying effects by species and PSP data.

Data processing

- ``PSPdataTypes`` -- which jurisdictional plot data to use, important because the default will attempt to use PSP data that may be inaccessible
- ``sppEquivCol`` -- the column name in ``sim$sppEquiv`` that ultimately determines which species

paramName	paramClass	default	min	max	paramDesc	cacheKey
-----------	------------	---------	-----	-----	-----------	----------

TABLE 4.2: List of *Biomass_speciesParameters* parameters and their description.

paramName	paramClass	default	min	max	paramDesc	cacheKey
balanceGrowth	logical	FALSE	NA	NA	If TRUE, give all fitted species the same growthcurve (sharedGrowthcurve) and set each species' mANPPproportion so that $K = \text{mANPPproportion} \times \text{maxB} \wedge (1 - \text{growthcurve})$ is the same for all (balanceGrowthK). Under LandR competition a small cohort grows with K, and any difference in K between species is winner-take-all, while the per-species fitted growthcurve values are statistically indistinguishable. maxB is the median over the species' rows of speciesEcoregion. growthcurve and mANPPproportion are then rounded to 3 decimals, not 2. Species that were not fitted are not changed. The fit of the new values to the PSP data is returned in speciesBalanceCheck, with a warning for species that are worse than their own best fit by more than 2 log-likelihood units. Default FALSE keeps the fitted values.	TRUE

paramName	paramClass	default	min	max	paramDesc	cacheKey
biomassModel	character	Lambert2005	NA	NA	The model used to calculate biomass from DBH. Can be either 'Lambert2005' or 'Ung2008'.	TRUE
landis	logical	FALSE	NA	NA	If TRUE, run in 'LANDIS mode': expose the fitted LANDIS-version growth curves (the scaled non-linear biomass-over-age curves shown in the LandR-vs-non-linear plot) per species as the output object speciesGrowthCurvesLandis, for use as inputs to LANDIS-II Biomass Succession. The per-species growth-curve parameters (growthcurve, mortalityshape, etc.) are written to species regardless. Default FALSE preserves the standard behaviour.	TRUE
maxBInFactorial	integer	5000	NA	NA	The arbitrary maximum biomass for the factorial simulations. This is a per-species maximum within a pixel	TRUE

178

4 LandR Biomass_speciesParameters Module

paramName	paramClass	default	min	max	paramDesc	cacheKey
excludeBECzonesHybridSpruce	character	BWBS, SWB	NA	NA	BC BEC zones whose PSPs are NOT relabelled by mergeHybridSprucePSP: the boreal zones, where 'Picea glauca' is true white spruce.	TRUE
mergeHybridSprucePSP	character	engelmann	NA	NA	Mirrors LandR.mergeHybridSpruce, and follows it by default. With 'engelmann', and the hybrid spruce merged into Pice_eng in sppEquiv, BC PSP records of 'Picea glauca' (how BC's interior hybrid-zone spruce is recorded) are relabelled 'Picea engelmannii x glauca' so they count as Pice_eng, except in the BEC zones of excludeBECzonesHybridSpruce. 'white' and NA do not relabel.	TRUE
minimumPlots	numeric	50	10	NA	Minimum number of PSP plots per species	TRUE
minDBH	integer	0	0	NA	Minimum diameter at breast height (DBH) in cm used to filter PSP data. Defaults to 0 cm, i.e. all tree measurements are used.	TRUE

paramName	paramClass	default	min	max	paramDesc	cacheKey
PSPdataTypes	character	all	NA	NA	Which PSP datasets to source, defaulting to all. Other available options include 'BC', 'AB', 'SK', 'ON', 'NB', 'NFI', and 'dummy'. 'dummy' should be used for unauthorized users.	TRUE
PSPperiod	numeric	1920, 2019	NA	NA	The years by which to subset sample plot data, if desired. Must be a vector of length 2	TRUE
quantileAgeSubset	numeric	99	1	100	Quantile by which to subset PSP data. As older stands are sparsely represented the oldest measurements become vastly more influential. This parameter accepts both a single value and a list of vectors, named according to sppEquivCol.	TRUE

180

4 LandR Biomass_speciesParameters Module

paramName	paramClass	default	min	max	paramDesc	cacheKey
speciesFittingApproach	character	focal	NA	NA	Either 'all', 'pairwise', 'focal' or 'single', indicating whether to pool all species into one fit, do pairwise species (for multiple cohort situations) do pairwise species, but using a focal species approach where all other species are pooled into 'other' or do one species at a time. If 'all', all species will have identical species-level traits.	TRUE
sppEquivCol	character	LandR	NA	NA	The column in <code>sim\$sppEquiv</code> data.table that defines individual species. The names should match those in the species table.	TRUE
standAgesForFitting	integer	21, 91	NA	NA	The minimum and maximum ages of the biomass-by-age curves used in fitting. It is generally recommended to keep this param under 200, given the low data availability of stands aged 200+, with some exceptions. For a closed interval, end with a 1, e.g. <code>c(31, 101)</code> .	TRUE

paramName	paramClass	default	min	max	paramDesc	cacheKey
sharedGrowthcurve	numeric	NA	NA	NA	Only used if balanceGrowth is TRUE. The growthcurve given to all fitted species. NA uses the median of the species' unrounded fitted growthcurve.	TRUE
targetK	numeric	NA	NA	NA	Only used if balanceGrowth is TRUE. The $K = \text{mANPPproportion} \times \text{maxB} \wedge (1 - \text{growthcurve})$ given to all fitted species. NA uses the median over species of K computed with sharedGrowthcurve and each species' fitted (unrounded) mANPPproportion.	TRUE
useHeight	logical	TRUE	NA	NA	Should height be used to calculate biomass (in addition to DBH). DBH is used by itself when height is missing.	TRUE
.plots	character	screen	NA	NA	Used by Plots function, which can be optionally used here	TRUE
.plotInitialTime	numeric	0	NA	NA	This describes the simulation time at which the first plot event should occur	TRUE
.plotInterval	numeric	NA	NA	NA	This describes the simulation time interval between plot events	TRUE

182

4 LandR Biomass_speciesParameters Module

paramName	paramClass	default	min	max	paramDesc	cacheKey
.saveInitialTime	numeric	NA	NA	NA	This describes the simulation time at which the first save event should occur	TRUE
.saveInterval	numeric	NA	NA	NA	This describes the simulation time interval between save events	TRUE
.studyAreaName	character	NA	NA	NA	Human-readable name for the growth curve filename. If NA, a hash of sppEquiv[[sppEquivCol]] will be used.	TRUE
.useCache	character	.inputOb...	NA	NA	Should this entire module be run with caching activated? This is generally intended for data-type modules, where stochasticity and time are not relevant	TRUE
.useParallel	integer	2	NA	NA	maximum number of threads/workers to use for data.table operations; passed to <code>data.table::setDTthreads</code> and should be ≤ 4 .	TRUE

4.2.4 List of outputs

The module produces the following outputs (Table 4.3). Note that `species` and `speciesEcoregion` are modified versions of the inputted objects with the same name.

Tables

- ``species`` and ``speciesEcoregion`` -- tables with calibrated trait values.
- ``speciesGAMMs`` -- the fitted GAMM model objects for each species.

TABLE 4.3: List of *Biomass_speciesParameters* output objects and their description.

objectName	objectClass	desc
<code>species</code>	<code>data.table</code>	The updated invariant species traits table (see above).
<code>speciesEcoregion</code>	<code>data.table</code>	The updated spatially-varying species traits table (see description for this object in inputs)
<code>speciesBalanceCheck</code>	<code>data.table</code>	An empty <code>data.table</code> unless <code>P(sim)\$balanceGrowth</code> is <code>TRUE</code> , when it has one row per fitted species: fitted and new <code>growthcurve</code> and <code>mANPPproportion</code> , the <code>maxB</code> used, the resulting <code>K</code> , and <code>deltaLL</code> , the log-likelihood units by which the new values fit the PSP data worse than the species' best fit (linearly interpolated on the factorial grid; NA outside the grid).
<code>speciesGrowthCurves</code>	<code>list</code>	list containing each species' non-linear model, model data, and the unfiltered PSP data

objectName	objectClass	desc
speciesGrowthCurvesLandis.data.table		An empty <code>data.table</code> unless <code>P(sim)\$landis</code> is TRUE, when it holds the fitted LANDIS-version growth curves (<code>BscaledNonLinear</code>) by species and standAge (the scaled non-linear curves shown in the <code>LandR_VS_NLM_growthCurves</code> plot), for use as LANDIS-II Biomass Succession inputs.
speciesGrowthCurvesPSPdata.table		An empty <code>data.table</code> unless <code>P(sim)\$landis</code> is TRUE, when it holds the PSP observations used to fit the growth curves (biomass by standAge and species, with <code>OrigPlotID1</code> for joining to plot locations / ecoregion), a diagnostic to plot against <code>speciesGrowthCurvesLandis</code> .

4.2.5 Simulation flow and module events

Biomass_speciesParameters initializes itself and prepares all inputs provided there is an active internet connection and the user has access to the data (and a Google Account to do so).

We advise future users to run *Biomass_speciesParameters* with defaults and inspect what the objects are like before supplying their own data. The user does not need to run *Biomass_speciesFactorial* to generate their own theoretical curves (unless they wish to), as the module accesses pre-generated theoretical curves.

Note that this module only runs once (in one “time step”) and only executes one event (`init`). The general flow of *Biomass_speciesParameters* processes is:

1. Preparation of all necessary data and input objects that do not require parameter fitting (e.g., the theoretical species growth curve data);

2. Sub-setting PSP data and calculating the observed species growth curves using non-linear growth models;
3. Finding the theoretical species growth curve that best matches the observed curve, for each species. Theoretical curves are subset to those with longevity matching the species' longevity (in species table) and with growthcurve and mortalityshape values;
4. Calibrating maxB and maxANPP.

4.3 Usage example

This module can be run stand-alone, but it won't do much more than calibrate species trait values based on dummy input trait values. We provide an example of this below, since it may be of value to run the module by itself to become acquainted with the calibration process and explore the fitted non-linear models. However, we remind that to run this example you will need a Google Account, and to be granted access to the data.

A realistic usage example of this module and a few others can be found in this repository¹⁰ and in Barros et al. [2].

4.3.1 Load Spades and other packages.

4.3.2 Set up R libraries

```
tempDir <- tempdir()

pkgPath <- file.path(tempDir, "packages", version$platform,
  paste0(version$major,
    ".", strsplit(version$minor, "[.]" )[[1]][1]))
dir.create(pkgPath, recursive = TRUE)
.libPaths(pkgPath, include.site = FALSE)
```

¹⁰https://github.com/CeresBarros/LandRBiomass_publication

```

repos <- c("predictiveecology.r-universe.dev",
getOption("repos"))
options(repos = repos)
install.packages("SpaDES.project") ## gets Require too

```

4.3.3 Get the module and module dependencies

```

library(Require)

paths <- list(inputPath = normPath(file.path(tempDir,
"inputs")),
  cachePath = normPath(file.path(tempDir, "cache")),
  modulePath = normPath(file.path(tempDir,
    "modules")), outputPath = normPath(file.path(tempDir,
    "outputs")))

SpaDES.project::getModule(modulePath = paths$modulePath,
c("PredictiveEcology/Biomass_speciesParameters@main"),
  overwrite = TRUE)
Require::Require("SpaDES.core (>= 2.1.4)")
Require("googledrive")

```

4.3.4 Setup simulation

```

times <- list(start = 0, end = 1)

modules <- list("Biomass_speciesParameters")

objects <- list()
inputs <- list()
outputs <- list()
parameters <- list()

```

```
mySim <- SpaDES.core::simInitAndSpades(times = times, params =
  parameters,
  modules = modules, paths = paths, objects = objects)

## to inspect the fitted growth models:
mySim$speciesGrowthCurves$Pice_mar
```

4.4 References

- [2] Ceres Barros et al. “Empowering ecological modellers with a PERFICT workflow: seamlessly linking data, parameterisation, prediction, validation and visualisation”. In: *Methods in Ecology and Evolution* 14.1 (2023), pp. 173–188. DOI: 10.1111/2041-210X.14034¹¹.
- [8] Dean W Coble and Young-Jin Lee. “Use of a generalized sigmoid growth function to predict site index for unmanaged loblolly and slash pine plantations in East Texas”. In: *Proceedings of the 13th Biennial Southern Silvicultural Research Conference*. Ed. by Kristina F. Connor. Gen. Tech. Rep. SRS-92. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station, 2006, pp. 291–295. URL: <https://research.fs.usda.gov/treesearch/23398>.
- [10] Desta Fekedulegn, Mairitin P. Mac Siurtain, and Jim J. Colbert. “Parameter estimation of nonlinear growth models in forestry”. In: *Silva Fennica* 33.4 (1999). Article id 653. DOI: 10.14214/sf.653¹².
- [16] M.-C. Lambert, C.-H. Ung, and F. Raulier. “Canadian national tree aboveground biomass equations”. In: *Canadian Journal of Forest Research* 35.8 (2005), pp. 1996–2018. DOI: 10.1139/x05-112¹³.
- [30] Chhun-Huor Ung, Pierre Bernier, and Xiao-Jing Guo. “Canadian national biomass equations: new parameter estimates that include British Columbia data”. In: *Canadian Journal of Forest Research* 38.5 (2008), pp. 1123–1132. DOI: 10.1139/X07-224¹⁴.

¹¹<https://doi.org/10.1111/2041-210X.14034>

¹²<https://doi.org/10.14214/sf.653>

¹³<https://doi.org/10.1139/x05-112>

¹⁴<https://doi.org/10.1139/X07-224>

LandR Validation Modules

LandR ‘validation modules’ differ from ‘data modules’ in that their objective is not to obtain input data and estimate parameters to run a simulation, but rather to confront simulation outputs against observed data – even if these modules can potentially obtain and pre-process the validation data. At the moment, only one validation module is available, *Biomass_validationKNN*, but we expect an increment in the number of validation modules as LandR usage expands.



5

LandR Biomass_validationKNN Module

Biomass validationKNN v. 1.0.1 ¹

Get help Report issues ²

5.0.0.1 Authors:

Ceres Barros ceres.barros@ubc.ca³ [aut, cre], Eliot McIntire eliot.mcintire@nrcan-rncan.gc.ca⁴ [aut]

This documentation is work in progress. Potential discrepancies and omissions may exist for the time being. If you find any, contact us using the “Get help” link above.

5.1 Module Overview

5.1.1 Quick links

- General functioning
- List of input objects
- List of parameters
- List of outputs
- Simulation flow and module events

¹https://github.com/PredictiveEcology/Biomass_validationKNN990f303a4e0b4559422455c7ef877acc333b87b0

²https://github.com/PredictiveEcology/Biomass_validationKNN/issues

³<mailto:ceres.barros@ubc.ca>

⁴<mailto:eliot.mcintire@nrcan-rncan.gc.ca>

5.1.2 Summary

LandR *Biomass_validationKNN* (hereafter *Biomass_validationKNN*) provides an approach to validate outputs from LandR Biomass (i.e., *Biomass_core* linked with other modules or not) simulations, using publicly available data for Canadian forests. It produces both a visual and statistical validation of *Biomass_core* outputs related to species abundance and presence/absence in the landscape. To do so, it downloads and prepares all necessary data (observed and simulated), calculates validation statistics and produces/saves validation plots.

5.1.3 Links to other modules

Biomass_validationKNN is intended to be used with *Biomass_core* and any other modules that link to it and affect cohort biomass (e.g., disturbance modules and calibration modules may both affect resulting biomass). See here⁵ for all available modules in the LandR ecosystem and select *Biomass_validationKNN* from the drop-down menu to see potential linkages. By default, disturbed pixels are excluded from the validation, but the user can bypass this option. The following is a list of the modules commonly validated with *Biomass_validationKNN*.

- *Biomass_core*⁶: core forest dynamics simulation module. Used downstream from *Biomass_borealDataPrep*;

Data and calibration modules:

- *Biomass_speciesData*⁷: grabs and merges several sources of species cover data, making species percent cover (% cover) layers used by other LandR Biomass modules. Default source data spans the entire Canadian territory;
- *Biomass_borealDataPrep*⁸: prepares all parameters and inputs (including initial landscape conditions) that *Biomass_core* needs to run a realistic simulation. Default values/inputs produced are relevant for boreal forests of Western Canada;

⁵https://rpubs.com/PredictiveEcology/LandR_Module_Ecosystem

⁶https://github.com/PredictiveEcology/Biomass_core

⁷https://github.com/PredictiveEcology/Biomass_speciesData

⁸https://github.com/PredictiveEcology/Biomass_borealDataPrep

- *Biomass_speciesParameters*⁹: calibrates four-species level traits using permanent sample plot data (i.e., repeated tree biomass measurements) across Western Canada.

Disturbance-related modules:

- *Biomass_regeneration*¹⁰: simulates cohort biomass responses to stand-replacing fires (as in the LANDIS-II Biomass Succession Extension v.3.2.1), including cohort mortality and regeneration through resprouting and/or serotiny;
- *Biomass_regenerationPM*¹¹: like *Biomass_regeneration*, but allowing partial mortality. Based on the LANDIS-II Dynamic Fuels & Fire System extension [29];
- *fireSense*: climate- and land-cover-sensitive fire model simulating fire ignition, escape and spread processes as a function of climate and land-cover. Includes built-in parameterisation of these processes using climate, land-cover, fire occurrence and fire perimeter data. Requires using *Biomass_regeneration* or *Biomass_regenerationPM*. See modules prefixed “fire-Sense_” at <https://github.com/PredictiveEcology/>;
- *LandMine*¹²: wildfire ignition and cover-sensitive wildfire spread model based on a fire return interval input. Requires using *Biomass_regeneration* or *Biomass_regenerationPM*;
- *scfm*¹³: spatially explicit fire spread module parameterised and modelled as a stochastic three-part process of ignition, escape, and spread. Requires using *Biomass_regeneration* or *Biomass_regenerationPM*.

⁹https://github.com/PredictiveEcology/Biomass_speciesParameters

¹⁰https://github.com/PredictiveEcology/Biomass_regeneration

¹¹https://github.com/PredictiveEcology/Biomass_regenerationPM

¹²<https://github.com/PredictiveEcology/LandMine>

¹³<https://github.com/PredictiveEcology/scfm>

5.2 Module manual

5.2.1 General functioning

Biomass_validationKNN compares simulated outputs of two years (across replicates), with corresponding years of observed data. It was designed to compare the observed data for years 2001 (start point for the simulation) and 2011 (i.e., after 10 years of simulation) of the kNN forest layers of the Canadian National Forest Inventory – these are currently the only available FAIR datasets [sensu 28] on stand biomass and species % cover changes across Canada. However, the user can supply other sources of observed data, as long as they have an identical format.

The validation is done both visually (using barplots and boxplots) and using two statistics: mean absolute deviation of simulated biomass (per species) and the sum of negative log-likelihoods (SNLL) of predictions with respect to observed data for species biomass, species presences/absences and changes in biomass (ΔB) – the later is still under development.

This module assumes that the simulation data preparation was carried out by *Biomass_borealDataPrep*, and so, to ensure that the comparison and the simulated datasets are built with the same assumptions, the data treatment steps in *Biomass_borealDataPrep* are repeated here.

The module may also exclude disturbed pixels identified in *rstLCCChange* raster layer and in the fire perimeter data (*firePerimeters* layer). If this is not intended, the user can provide a *rstLCCChange* with NA's only and/or an empty *firePerimeters* sf object.

Biomass_validationKNN then compares simulated species biomass, presences, dominance, and changes in biomass against observed data available for the starting conditions (2011 by default) and for second time point (e.g. 2011, or after 10 years of simulation). To do so, for each year and replicate, and for both the simulated and observed data, the module calculates:

- species relative abundances at the pixel- and landscape-level (across all pixels);
- species presences and dominance at the landscape level;
- changes in species biomass (ΔB) at the pixel- and landscape-level for both

the simulated and observed data. Biomass units respect those used in *Biomass_core* (g/m^2).

Pixel-level relative abundances are calculated as the species biomass (summed across cohorts) divided by the total pixel biomass (summed across cohorts and species), while landscape-wide relative abundances are calculated as the sum of a species biomass across all pixels divided by the sum of total biomass across all pixels.

Species presences are calculated as the number of pixels where a given species is present, and species dominance is calculated as the number of pixels where a species has the highest relative biomass in a given pixel. Pixels where two or more species share the highest biomass value are classified as 'mixed forest', and pixels without any biomass are classified as 'no veg'.

Finally, ΔB is calculated per species as the final biomass (e.g., year 2011) minus the initial biomass (e.g., year 2001), either at the pixel- or landscape-level.

5.2.2 Validation approaches

5.2.2.1 Visual validation

The module plots the above metrics as barplots showing landscape-level values (averaged across replicates for the simulated data) or boxplots showing pixel-level values. Plotting can be live and/or in the form of exported images (or both turned off completely).

5.2.2.2 Mean absolute deviation

Mean absolute deviance (MAD) values are calculated on landscape- and pixel-level species relative abundances and ΔB , and landscape-level species presences and dominance. MAD values are calculated per replicate and year, except ΔB MAD values, which integrate across years. Output tables with MAD values are exported as `landscapeMAD` and `pixelMAD`, and the module also produces visual inspection of these values as dot-and-whisker plots.

5.2.2.3 Sum of negative log-likelihood (SNLL)

To provide a measure of overall goodness of fit of the model set-up that gave rise to the outputs, this is the set of starting conditions, parameters

and simulation mechanisms that generated predictions (which includes the LandR modules used), *Biomass_validationKNN* estimates sum of negative log-likelihoods (SNLL) of species presences (at the landscape-level), simulated species biomasses, and ΔB (the latter two at the landscape and pixel levels), with respect to their observed counterparts.

More precisely, let ℓ be the log-likelihood function denoting the probability of observing x of X (a random variable following a continuous probability distribution $f(x)$), given a parameter θ :

$$\ell(\theta | x) = f(x) \quad (5.1)$$

In our case, θ is equivalent to the model's starting conditions and structure, X is the observed data with x being the simulated values, and $f(x)$ the continuous probability distribution of X . For each variable that we wanted to evaluate and for each simulation replicate, Equation (5.1) is applied to calculate the SNLL estimated for each value of x at the pixel or landscape-level, i :

$$-\sum_{i=1}^N \ell(\theta | x_i) \quad (5.2)$$

where N is equal to total number of pixels. At the landscape scale $N = 1$.

For species presences and species biomass, we draw the probability of observing x_i (a vector of species presences/biomasses in pixel/landscape i) from a multinomial density distribution ($f(x_i) = \text{Multi}(n_i, p_i)$), where $n_i = \sum_{j=1}^K X_{i,j}$ (X being the observed values of biomass of $j = 1, \dots, K$ species in a pixel/landscape i) and p_i is the vector of simulated values $x_{i,j}$.

The computation of SNLL for ΔB is still under development. The following approach is currently implemented, but presents issues:

For ΔB , we draw the probability of observing $x_{i,j}$ (the simulated ΔB of $j = 1, \dots, K$ species in a pixel/landscape i) from a multivariate Gaussian distribution, $f(x_i) = \mathcal{N}(\mu_i, M_i)$, where μ_i is the vector of observed mean ΔB for each species $j = 1, \dots, K$, and M is the observed $K * K$ variance-covariance matrix of species ΔB . Unfortunately this is presenting problems, due to M not being strictly positive definite.

After calculating SNLL across pixels (or for the entire landscape), values are

averaged across replicates for an overall model estimate and exported in the `logLikelihood` table.

We refer to the Wikipedia pages on the multinomial distribution¹⁴ and on the multivariate Gaussian distribution¹⁵ for a good summary of these two distributions and their use in SNLL estimation.

5.2.3 List of input objects

The full list of input objects *Biomass_validationKNN* requires is presented below (Table 5.1). All have defaults except `studyArea`, which **must** be provided by the user, or another module.

Of these, the input spatial layers land-cover change (change type and year), fire perimeters, % species cover, stand age and stand biomass are obtained from National Forest Inventory kNN layers for years 2001 and 2011.

We strongly recommend that for the “starting point layers” (those suffixed with `*Start`, which by default correspond to 2001) the user supplies the same objects used as the starting input layers to initialise the simulation to make sure that they match.

Note that objects suffixed with `*Start` correspond to the same objects in the main simulation without this suffix (e.g. `rawBiomassMapStart` is `rawBiomassMap` in *Biomass_borealDataPrep*), whereas other objects like `studyArea` and `rasterToMatch` have the same names in the simulation and should be **exactly** the same object.

Of the inputs in Table 5.1, the following deserve special attention:

Spatial layers

- `biomassMap` – a map of simulated stand biomass (in g/m^2) filtered for the pixels where cohort dynamics were simulated. This corresponds to the `sim$biomassMap` object produced by *Biomass_borealDataPrep* or to the `sim$simulatedBiomassMap` produced by *Biomass_core*.
- `firePerimeters` – a fire perimeters polygon map that should be used to exclude recently burned pixels from the analysis. If this is not desired the user needs to provide an empty `sf` object (e.g., `sf::st_polygon()`).

¹⁴https://en.wikipedia.org/wiki/Multinomial_distribution

¹⁵https://en.wikipedia.org/wiki/Multivariate_normal_distribution#Density_function

- `rawBiomassMapStart` – raw biomass data used to initialise and parametrise `Biomass_core`. By default, the module uses the stand biomass map from kNN for the year 2001. The user must make sure this is appropriate for their use case, or else supply the correct raster layer.
- `rawBiomassMapEnd` – raw biomass data used to validate the model after several simulation years. By default, the module uses the kNN stand biomass map from 2011, which is compared with the 10th year of a simulation initialised using the kNN 2001 data. The user must make sure this is appropriate for their use case, or else supply the correct raster layer.
- `rstLCChange` – a binary raster layer with disturbed pixels that should be removed from the analyses. Can be combined with `rstLCChangeYr` to filter pixels disturbed in a given time period defined by `P(sim)$LCChangeYr`. Defaults to Canada's forest change national map between 1985-2011 (CFS)¹⁶.
- `rstLCChangeYr` – a raster layer with year of disturbance. This is an optional layer that can be combined with `rstLCChange` and `P(sim)$LCChangeYr` to filter disturbed pixels by year of disturbance. Not used by default. Defaults to Canada's forest change year national map between 1985-2011 (CFS)¹⁷.
- `speciesLayersStart` – same as `rawBiomassMapStart`, but with respect to species % cover data.
- `speciesLayersEnd` – same as `rawBiomassMapEnd`, but with respect to species % cover data.
- `studyArea` – shapefile. A `SpatialPolygonsDataFrame` with a single polygon determining where the simulation will take place. This is the only input object that **must be supplied by the user**.

Simulation-related objects

- `allCohortData` – OPTIONAL. A `data.table` containing all `cohortData` objects relevant for the validation (e.g., as many `cohortData` objects as simulation replicates times 2, for the beginning and end year). If not supplied, `Biomass_validationKNN` attempts to produce this object using the `cohortData` object file listed in `simulationOutputs`. Hence, the user must either supply *both* `allCohortData` and `pixelGroupMapStk` or `simulationOutputs`.

¹⁶https://opendata.nfis.org/downloads/forest_change/C2C_change_type_1985_2011.zip

¹⁷https://opendata.nfis.org/downloads/forest_change/C2C_change_year_1985_2011.zip

- `pixelGroupMapStk` – OPTIONAL. As `allCohortData`, but with respect to `pixelGroupMap` objects.
- `simulationOutputs` – OPTIONAL. A `data.frame` that has the same structure as the `data.frame`'s specifying outputs to be saved in `spades(..., outputs = data.frame(...))`. We advise passing the same `data.frame` that was supplied to `spades` during the simulation call, but filtered by the relevant `cohortData` and `pixelGroupMap` objects and, potentially, with file paths corrected to match the current working directory (see [Usage example](#)). Only used if `allCohortData` and `pixelGroupMapStk` are not supplied.
- `pixelGroupMap` – a raster layer with *pixelGroup* IDs per pixel. Pixels are grouped based on identical *ecoregionGroup*, *speciesCode*, *age* and *B* composition, even if the user supplies other initial groupings (e.g., this is possible in the *Biomass_borealDataPrep* data module).

TABLE 5.1: List of *Biomass_validationKNN* input objects and their description.

objectName	objectClass	desc	sourceURL
allCohortData	data.table	A table 'rbinding' all cohortData tables saved during the simulation, with the added 'rep' and 'year' columns. It must contain all the validation years. If not supplied, the module will attempt to retrieve them using the 'simulationOutputs' table	NA
biomassMap	SpatRaster	total biomass raster layer in study area (in g/m ²), filtered for pixels covered by cohortData. Only used to calculate total no. of pixels being simulated If not supplied, will default to rawBiomassMapStart	NA
firePerimeters	sf	A map of fire perimeters in the study area that can be used to exclude pixels that have been burnt during the validation period. If burnt pixels are not to be excluded Provide an empty sf object with the same properties as the default. Defaults to the latest Canadian Wildland Fire Information System National Burned Area Composite, subset to fires occurring up to last validation year (inclusively) . Source URL determined by fireURL	NA

objectName	objectClass	desc	sourceURL
fireURL	character	A URL to a fire database, such as the Canadian National Fire Database, that is a zipped shapefile with fire polygons, an attribute (i.e., a column) named 'Year'. If supplied (omitted with NULL or NA), this will be used to 'update' age pixels on standAgeMap with 'time since fire' as derived from this fire polygons map	https://cwfis.cfs.nrcan.gc.ca/downloads/nfdb/fire_poly/current_version/NFDB_poly.zip
pixelGroupMapStk	SpatRaster	A stack of pixelGroupMaps saved during the simulation, particularly for the validation years. If not supplied, the module will attempt to make it using the 'simulationOutputs' table. names(pixelGroupMapStk) should be 'paste0('year', , '_rep',)'.	NA

objectName	objectClass	desc	sourceURL
rawBiomass MapStart	SpatRaster	observed total biomass raster layer in study area at the first year of the validation period. Defaults to the Canadian Forestry Service, National Forest Inventory, kNN-derived total aboveground biomass map from 2001 (in ton/ha). See https://open.canada.ca/data/en/dataset/ec9e2659-1c29-4ddb-87a2-6aced147a990 for metadata.	http://ftp.maps.canada.ca/pub/nrcan_rncan/Forests_Foret/canada-forests-attribute_s_attributes-forests-canada/2001-attributes_2001/NFI_MODIS250m_2001_kNN_StructureBiomass_TotalLiveAboveGround v1.tif

objectName	objectClass	desc	sourceURL
rawBiomass MapEnd	SpatRaster	observed total biomass raster layer in study area at the last year of the validation period. Defaults to the Canadian Forestry Service, National Forest Inventory, kNN-derived total aboveground biomass map from 2011 (in ton/ha) See https://open.canada.ca/data/en/dataset/ec9e2659-1c29-4ddb-87a2-6aced147a990	http://ftp.mmaps.canada.ca/pub/nrcan_rncan/Forests_Foret/canada-forests-attribute_s_attribute_s-forests-canada/2011-attribute_s_attribute_s-2011/NFI_MODIS250m_2011_kNN_StructureBiomass_TotalLiveAboveGround.v1.tif
rasterToMatch	SpatRaster	A raster of the studyArea in the same resolution and projection as rawBiomass_MapStart. This is the scale used for all outputs for use in the simulation.	NA

objectName	objectClass	desc	sourceURL
rstLCChange	SpatRaster	A mask-type map of land cover changes in the study area that can be used to exclude pixels that have been disturbed during the validation period. If disturbed pixels are not to be excluded Provide an empty sf object with the same properties as the default. Defaults to Canada's forest change map between 1985-2011 (CFS), filtered for years 2001-2011 (inclusive ly) and all disturbances collapsed (map only has values of 1 and NA). See <code>P(sim)\$LCChangeYr</code> parameter to change the period of disturbances, and https://openda.ta.nfis.org/mapserver/nfis-change_eng.html for more information.	https://openda.ta.nfis.org/mapserver/nfis-change_eng.html
rstLCChangeYr	SpatRaster	An OPTIONAL map of land cover change years in the study area used to exclude pixels that have been disturbed during the validation period. It defaults to Canada's forest change year national map between 1985-2011 (CFS). If <code>P(sim)\$LCChangeYr</code> is not NULL, this layer is used to filter disturbed pixels that fall within the years specified by <code>P(sim)\$LCChangeYr</code> . If <code>P(sim)\$LCChangeYr</code> is NULL this layer is not used. See https://openda.ta.nfis.org/mapserver/nfis-change_eng.html for more information.	https://openda.ta.nfis.org/mapserver/nfis-change_eng.html

objectName	objectClass	desc	sourceURL
simulation Outputs	data.table	An OPTIONAL table listing simulation out puts (as passed to <code>spades()</code> , or <code>experiment</code>) that will be used to make <code>allCo hortData</code> , <code>pixelGroupMapStk</code> , if these are not provided.	NA
speciesLay ersStart	RasterStack	observed cover percentage raster layers by species in Canada species map, at the first year of the validation period. Defaults to the Canadian Forestry Service, National Forest Inventory, kNN-derived species cover maps from 2001, using a cover threshold of 10% - see https://open.canada.ca/data/en/dataset/ec9e2659-1c29-4ddb-87a2-6aced147a990 for metadata	http://ftp.maps.canada.ca/pub/nrcan_rncan/Forests_Foret/canada-forests-attribute_s_attribute_s-forests-canada/2001-attribute_s_attribute_s-2001/

objectName	objectClass	desc	sourceURL
speciesLayersEnd	RasterStack	observed percent cover raster layers by species in Canada used for validation at the last year of the validation period. Defaults to the Canadian Forestry Service, National Forest Inventory, kNN-derived species cover maps from 2011 - see https://open.canada.ca/data/en/dataset/ec9e2659-1c29-4ddb-87a2-6aced147a990 for metadata	http://ftp.mmaps.canada.ca/pub/nrcan_rncan/Forests_Foret/canada-forests-attribute_s_attributes-forests-canada/2011-attributes-2011/
sppColorVect	character	A named vector of colors to use for plotting. The names must be in <code>simSppEquiv[[simSppEquivCol]]</code> , and should also contain a color for 'Mixed'	NA
sppEquiv	data.table	table of species equivalencies. See <code>LandR::sppEquivalencies_CA</code> .	NA

objectName	objectClass	desc	sourceURL
standAgeMapStart	SpatRaster	observed stand age map in study area, at the first year of the validation period Defaults to the Canadian Forestry Service, National Forest Inventory, kNN-derived biomass map from 2001 - see https://open.canada.ca/data/en/dataset/ec9e2659-1c29-4ddb-87a2-6aced147a990 for metadata	http://ftp.maps.canada.ca/pub/nrcan_rncan/Forests_Foret/canada-forests-attribute_s_attributes-forests-canada/2001-attributes_attributes-2001/NFI_MODIS250m_2001_kNN_Structure_Stand_Age_v1.tif

objectName	objectClass	desc	sourceURL
standAgeMapEnd	SpatRaster	observed stand age raster layer in study area, at the last year of the validation period. Defaults to the Canadian Forestry Service, National Forest Inventory, kNN-derived stand age map from 2011. See https://open.canada.ca/data/en/dataset/ec9e2659-1c29-4ddb-87a2-6aced147a990	http://ftp.maps.canada.ca/pub/nrcan_rncan/Forests_Foret/canada-forests-attribute_s_attributes-forests-canada/2011-attributes_attributes-2011/NFI_MODIS250m_2011_kNN_Structure_Stand_Age_v1.tif
studyArea	SpatVector	Polygon to use as the study area. Must be provided by the user	NA

5.2.4 List of parameters

Table 5.2 lists all parameters used in *Biomass_validationKNN* and their detailed information. All have default values specified in the module's meta-data

Of the parameters listed in Table 5.2, the following are particularly important:

- `LCChangeYr` – integer. Optional parameter defining the years of disturbance that should be filtered out of the analysis using the `rstLCChangeYr` layer. This parameter is set to `NULL` by default, meaning that `rstLCChangeYr` will not be used.
- `sppEquivCol` – character. the column name in `speciesEquivalency` data.table that defines the naming convention to use throughout the simulation.
- `validationReps` – integer. which simulation replicates should be used for the validation.
- `validationYears` – integer. What simulation years should be used for the validation - the year number needs to match the observed data year. For instance, if the first observed data year is 2001, that must be the first simulation year.

TABLE 5.2: List of *Biomass_validationKNN* parameters and their description.

paramName	paramClass	default	min	max	paramDesc	cacheKey
coverThresh	integer	10	NA	NA	The minimum % cover a species needs to have (per pixel) in the study area to be considered present. Should be the same as the one used to obtain the species cover layers for simulation set up.	TRUE
deciduousCoverDiscount	numeric	0.8418911	NA	NA	This was estimated with data from NWT on March 18, 2020 and may or may not be universal. Should be the same as the one used when preparing cohortData in the simulation set up.	TRUE

paramName	paramClass	default	min	max	paramDesc	cacheKey
LCChangeYr	integer		1900	NA	OPTIONAL. An integer or vector of integers of the validation period years, defining which years of land-cover changes (i.e. disturbances) should be excluded. <code>NULL</code> by default, which presumes no subsetting based on years is done internally (either the user supplies a pre-filtered <code>rstLCChange</code> , or no filtering is desired). If not <code>NULL</code> <code>rstLCChangeYr</code> is used to filter disturbed pixels within the specified years. See https://opendata.nfis.org/mapserver/nfis-change_eng.html for more information.	TRUE
minCoverThreshold	numeric	5	0	100	Cover that is equal to or below this number will be omitted from the dataset Should be the same as the one used when preparing <code>cohortData</code> in the simulation set up.	TRUE
obsDeltaAgeB	logical	TRUE	NA	NA	When TRUE, the observed changes in biomass and age (<code>deltaB</code> , <code>deltaAge</code>) between the two validation years will be plotted as maps and scatterplots	TRUE

paramName	paramClass	default	min	max	paramDesc	cacheKey
pixelGroupBiomassClass	numeric	100	NA	NA	When assigning pixelGroup membership, this defines the resolution of biomass that will be considered 'the same pixelGroup', e.g., if it is 100, then 5160 and 5240 will be the same. Should be the same as the one used when preparing cohortData in the simulation set up.	TRUE
sppEquivCol	character	Boreal	NA	NA	The column in sim\$sppEquiv data.table to use as a naming convention	TRUE
validationReps	integer	1, 2, 3,....	NA	NA	The simulation repetitions for the validation. Defaults to 1:10. Set to NA if not using repetitions (i.e. only one run)	TRUE
validationYears	integer	2001, 2011	NA	NA	The simulation years for the validation. Defaults to 2001 and 2011. Must select two years	TRUE
.plotInitialTime	integer	1	NA	NA	If NA plotting is off completely (this includes saving).	TRUE
.plots	character	object, png	NA	NA	Passed to types in Plots (see ?Plots). There are a few plots that are made within this module, if set. Note that plots (or their data) are saved in file.path(outputPath(sim), 'figures'). If NA, plotting is off completely (this includes plot saving).	TRUE

212

5 LandR Biomass_validationKNN Module

paramName	paramClass	default	min	max	paramDesc	cacheKey
.saveInitialTime	numeric	NA	NA	NA	This describes the simulation time at which the first save event should occur	TRUE
.saveInterval	numeric	NA	NA	NA	This describes the simulation time interval between save events	TRUE
.sslVerify	integer	64	NA	NA	Passed to <code>httr::config(ssl_verifypeer = P(sim)\$sslVerify)</code> when downloading KNN (NFI) datasets. Set to 0L if necessary to bypass checking the SSL certificate (this may be necessary when NFI's website SSL certificate is not correctly configured).	TRUE
.studyAreaName	character	NA	NA	NA	Human-readable name for the study area used. If NA, a hash of <code>studyArea</code> will be used.	TRUE
.useCache	logical	.inputOb....	NA	NA	Controls cache; caches the init event by default	TRUE

5.2.5 List of outputs

The module produces the following outputs (Table 5.3):

TABLE 5.3: List of *Biomass_validationKNN* output objects and their description.

objectName	objectClass	desc
logLikelihood	data.table	A table of negative sum log-likelihood values calculated for different variables and averaged across repetitions. At the moment, log-likelihood values are calculated for biomass (landscape- and pixel-level), species presences and dominance (landscape-level) and deltaB (landscape- and pixel-level. For biomass and count data (presences/dominance, we assume an underlying multinomial distribution, and for deltaB a multivariate Gaussian distribution - note that the later is still under development.
landscapeMAD	data.table	Mean absolute deviance values calculated on landscape-level relative abundances, species presences and dominance, and deltaB, per repetition and year (except for deltaB, which is integrated across years)

objectName	objectClass	desc
landscapeVars	data.table	A table containing observed and simulated landscape-averaged variables used for validation (by year and repetition, 'rep', in the case of simulated data), namely: species relative abundances ('relAbund'), species presences ('count'), species dominance (as in no. pixels where a given species, has higher 'relAbund'; 'countDom') and species changes in biomass, as 2011 minus 2001 ('deltaB'). Observed data rows are labelled as 'observed' in 'dataType' column. In species dominance, pixels with ≥ 2 species with max(B) and pixels with no B are classified as 'Mixed' and 'No veg.', respectively in the 'speciesCode' column - note that this is 'vegType' column in pixelCohortData.
pixelCohortData	data.table	A table containing observed and simulated pixel-level data (by year and repetition, 'rep', in the case of simulated data) on species biomass (summed across cohorts, 'B'), total pixel biomass ('pixelB'), average biomass-weighted pixel age ('pixelAge'), species relative abundance (calculated as B/pixelB , 'relativeAbund'), species dominance (the species with max(B), 'vegType'), and landscape-wide biomass ('landscapeB'). Observed data columns are suffixed with 'Obsrvd'. In species dominance, pixels with ≥ 2 species with max(B) (i.e. 'noDoms' ≥ 2) are classified as 'Mixed'.

objectName	objectClass	desc
pixelMAD	data.table	Mean absolute deviance values calculated on pixel-level relative abundances and deltaB, per repetition and year (except for deltaB, which is integrated across years)
pixelVars	data.table	The same as landscapeVars, but variables are calculated at the pixel-level
rstDisturbedPix	SpatRaster	Raster of pixel IDs (as a mask) that have been disturbed by fire or suffered land-cover changes during the validation period. These pixels are excluded from the validation.
rawBiomassMapStart	SpatRaster	observed total biomass raster layer in study area at the first year of the validation period. Filtered to exclude pixels that were disturbed during the validation period
rawBiomassMapEnd	SpatRaster	observed total biomass raster layer in study area at the last year of the validation period. Filtered to exclude pixels that were disturbed during the validation period
speciesLayersStart	SpatRaster	observed percent cover raster layers by species in Canada at the first year of the validation period. Filtered to exclude pixels that were disturbed during the validation period
speciesLayersEnd	SpatRaster	observed percent cover raster layers by species in Canada at the last year of the validation period. Filtered to exclude pixels that were disturbed during the validation period
standAgeMapStart	SpatRaster	observed stand age map in study area, at the first year of the validation period Filtered to exclude pixels that were disturbed during the validation period

objectName	objectClass	desc
standAgeMapEnd	SpatRaster	observed stand age map in study area, at the last year of the validation period Filtered to exclude pixels that were disturbed during the validation period

5.2.6 Simulation flow and module events

Biomass_validationKNN initialises itself and prepares all inputs provided that it has access to outputs of simulations from *Biomass_core*, and internet access to retrieve the observed kNN datasets used for validation¹⁸.

The module then compiles all simulation output data provided that the user supplies the object names and their file paths via the `simulationOutputs` input object. Alternatively, the user may pass the pre-compiled outputs (namely the `cohortData` and `pixelGroupMap` objects) via the `allCohortData` and `pixelGroupMapStk` input objects. See [list of input objects](#) for more detail.

Future users should run *Biomass_validationKNN* with defaults and inspect what the objects are like before supplying their own data, or alternative data URLs. Alternatively, users may develop their own validation modules using *Biomass_validationKNN* as a template. We expect the number of validation modules to increase as other validation approaches are developed based on project needs.

The general flow of *Biomass_validationKNN* processes is (note that this module only runs once, i.e. in one “time step”):

1. Preparation of all necessary objects, namely obtaining the observed data layers from on-line repositories (or if available stored local copies) and the compiling simulated data if the user has not done so previously (see [list of input objects](#)) – (`init` event).
2. Calculation of summary variables for validation (`calculateValidVars` event), namely :
 - relative biomass per species per pixel and across the landscape (per year and per replicate)
 - changes in species biomass per pixel and across the landscape (per replicate), with respect to the first year.
 - species dominance across the landscape
 - species presences across the landscape
3. Calculation of validation statistics (`validationStats` event),

¹⁸Raw data layers downloaded by the module are saved in ‘`dataPath(sim)`’, which can be controlled via ‘`options(reproducible.destinationPath = ...)`’.

namely mean absolute deviations (MAD) and sum of negative log-likelihoods (SNLL).

4. Assessment of the relationship between observed ΔB and observed ΔAge (`obsDeltaMaps` event) – this is an optional visual diagnostic of the observed data that produces scatterplots of $\Delta B \sim \Delta \text{Age}$ of three types:

- With raw observed values of ΔB and ΔAge
- With ΔB and ΔAge calculated on observed data *after* pre-processing (i.e., the data clean-steps done in `Biomass_borealDataPrep`, which are also done to the observed data before validation)
- With the data shown in 2) above, but filtered by pixels where there was only a stand age increment corresponding to the number of years of between the two validation time points. This is not necessarily a *correct* filter, as stands may have suffered an age reduction due to the loss of old cohorts from background mortality (i.e., not coming from disturbances. However, if using the default input datasets, it is unlikely that this is a widespread phenomenon in only 10 years. We remind the user that disturbed pixels should be removed from the analyses when validating succession dynamics in the absence of disturbance - the default option.

5. Plots (`landscapeWidePlots`, `pixelLevelPlots` and `deltaBComparisons` events):

- Barplots of landscape-wide and pixel-level comparisons between observed and simulated data, with respect to relative biomass, dominance and presences.
- Boxplots of biomass changes (ΔB) in observed and simulated data, with respect to the first year.
- Maps of biomass and age changes (ΔB , ΔAge) with respect to the first year, in observed and simulated data.

All module default outputs are in the form of plots, but the user can chose to save any objects (see Table 5.3).

5.3 Usage example

5.3.1 Set up R libraries

```
options(repos = c(CRAN = "https://cloud.r-project.org"))
tempDir <- tempdir()

pkgPath <- file.path(tempDir, "packages", version$platform,
  paste0(version$major,
    ".", strsplit(version$minor, "[.]")[[1]][1]))
dir.create(pkgPath, recursive = TRUE)
.libPaths(pkgPath, include.site = FALSE)

if (!require(Require, lib.loc = pkgPath)) {
  remotes::install_github(paste0("PredictiveEcology/",
    "Require@5c44205bf407f613f53546be652a438ef1248147"),
    upgrade = FALSE, force = TRUE)
  library(Require, lib.loc = pkgPath)
}

setLinuxBinaryRepo()
```

5.3.2 Get the module and module dependencies

Because *Biomass_validationKNN* is meant to validate simulation outputs against observed data, we need to first run a simulation of forest dynamics with *Biomass_core*. To do that we get both modules' code from the PredictiveEcology GitHub repository (all install all necessary packages). Notice that we are placing all packages, module code, inputs and outputs in temporary directories.

```
Require(paste0("PredictiveEcology/",
  "SpaDES.project@6d7de6ee12fc967c7c60de44f1aa3b04e6eeb5db"),
  require = FALSE, upgrade = FALSE, standAlone = TRUE)
```

```

paths <- list(inputPath = normPath(file.path(tempDir,
"inputs")),
  cachePath = normPath(file.path(tempDir, "cache")),
  modulePath = normPath(file.path(tempDir,
    "modules")), outputPath = normPath(file.path(tempDir,
    "outputs")))

SpaDES.project::getModule(modulePath = paths$modulePath,
c("PredictiveEcology/Biomass_core@master",
  "PredictiveEcology/Biomass_validationKNN@master"), overwrite
= TRUE)

## make sure all necessary packages are installed:
outs <- SpaDES.project::packagesInModules(modulePath =
paths$modulePath)
Require(c(unname(unlist(outs)),
"PredictiveEcology/SpaDES.experiment@development",
  "future"), require = FALSE, standAlone = TRUE)

## load necessary packages
Require(c("SpaDES.core", "LandR", "reproducible", "pemisc",
"SpaDES.experiment",
  "future"), upgrade = FALSE, install = FALSE)

```

5.3.3 Setup simulation

```

times <- list(start = 2001, end = 2011)

studyArea <- Cache(randomStudyArea, size = 1e+07) # cache this
so it creates a random one only once on a machine

# Pick the species you want to work with -- using the
# naming convention in 'Boreal' column of
# LandR::sppEquivalencies_CA

```

```

speciesNameConvention <- "Boreal"
speciesToUse <- c("Pice_Gla", "Popu_Tre", "Pinu_Con")

sppEquiv <- sppEquivalencies_CA[get(speciesNameConvention) %in%
  speciesToUse]
# Assign a colour convention for graphics for each species
sppColorVect <- sppColors(sppEquiv, speciesNameConvention,
  newVals = "Mixed",
  palette = "Set1")

## Usage example
modules <- as.list("Biomass_core")
objects <- list(studyArea = studyArea, sppEquiv = sppEquiv,
  sppColorVect = sppColorVect)

successionTimestep <- 20L

## keep default values for most parameters (omitted from
## this list)
parameters <- list(Biomass_core = list(sppEquivCol =
  speciesNameConvention,
  successionTimestep = successionTimestep, .plotInitialTime =
  times$start,
  .plotInterval = 1L, .plots = "png", .saveInitialTime =
  times$start,
  .useCache = "init", .useParallel = FALSE))

outputs <- data.frame(expand.grid(objectName = "cohortData",
  saveTime = unique(seq(times$start, times$end, by = 1)),
  eventPriority = 1,
  stringsAsFactors = FALSE))
outputs <- rbind(outputs, data.frame(objectName =
  "pixelGroupMap",
  saveTime = unique(seq(times$start, times$end, by = 1)),
  eventPriority = 1))

```

5.3.4 Run simulation

Here we run a simulation with three replicates using the `experiment2` function of the `SpaDES.experiment` R package [18], which builds a folder structure where simulation outputs are conveniently organised.

```
opts <- options(reproducible.useCache = TRUE,
  reproducible.destinationPath = paths$inputPath,
  spades.useRequire = FALSE)
graphics.off()
mySimInit <- simInit(times = times, params = parameters, modules
= modules,
  objects = objects, paths = paths, outputs = outputs)

plan(sequential)
mySimExperiment <- experiment2(sim1 = mySimInit, clearSimEnv =
FALSE,
  replicates = 3)
```

5.3.5 Validate simulation outputs with *Biomass_validationKNN*

Note that because we ran *Biomass_core* by itself using theoretical input data, we can expect the validation to reveal that the module didn't do a great job at reproducing observed patterns.

Note that a Google Account is needed to download some of the input files.

```
simulationOutputs <- lapply(mySimExperiment, FUN = function(x,
  localSimPaths) {
  oldPath <- dirname(outputPath(x)) ## exclude sim*_rep*
  folder
  DT <- as.data.table(outputs(x))
  DT[, file := sub(oldPath, localSimPaths$outputPath, file)]
  DT
}, localSimPaths = as.list(normPath(paths)))
simulationOutputs <- rbindlist(simulationOutputs)
```

```

validationPaths <- as.list(normPath(paths))
validationPaths$outputPath <-
file.path(validationPaths$outputPath,
           "validation")

validationTimes <- list(start = 1, end = 1)
validationParams <- list(Biomass_validationKNN =
list(sppEquivCol = params(mySimInit)$Biomass_core$sppEquivCol,
      validationReps = as.integer(1:3) ## or length of simLists
,
      validationYears = as.integer(c(2001, 2011)), .plots =
c("png")))

## make an empty fire polygon object to bypass removing
## fire-disturbed pixels
noFires <- sf::st_polygon()
validationObjects <- list(biomassMap =
mySimExperiment$sim1_rep1$biomassMap,
  firePerimeters = noFires, rasterToMatch =
mySimExperiment$sim1_rep1$rasterToMatch,
  rawBiomassMapStart = mySimExperiment$sim1_rep1$biomassMap,
  simulationOutputs = simulationOutputs, speciesLayersStart =
mySimExperiment$sim1_rep1$speciesLayers,
  sppColorVect = mySimExperiment$sim1_rep1$sppColorVect,
  sppEquiv = mySimExperiment$sim1_rep1$sppEquiv,
  studyArea = mySimExperiment$sim1_rep1$studyArea)

mySimValidation <- simInitAndSpades(times = validationTimes,
  params = validationParams, modules =
"Biomass_validationKNN",
  objects = validationObjects, paths = validationPaths,
  .studyAreaName = SName)

```

Here are some of the output figures automatically produced by *Biomass_validationKNN*

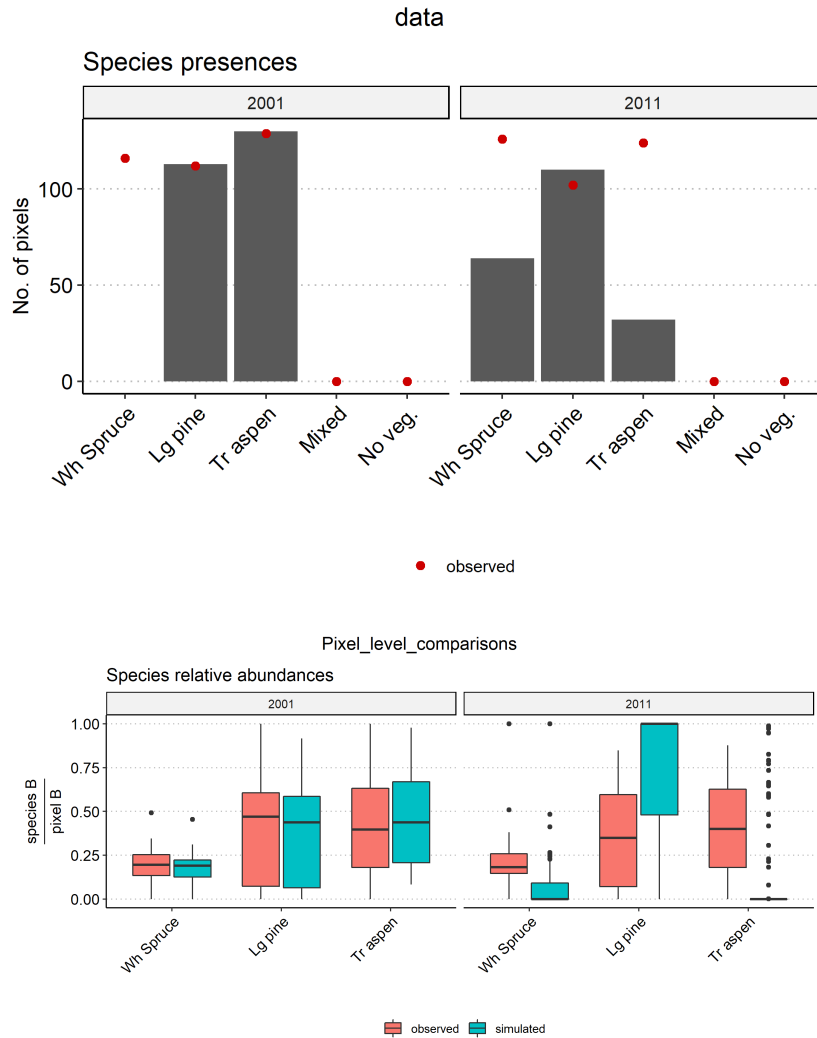


FIGURE 5.1: *Biomass_validationKNN* automatically generates plots showing a visual comparison between simulated and observed species presences (right) across the landscape, and relative species biomass per pixel (left).

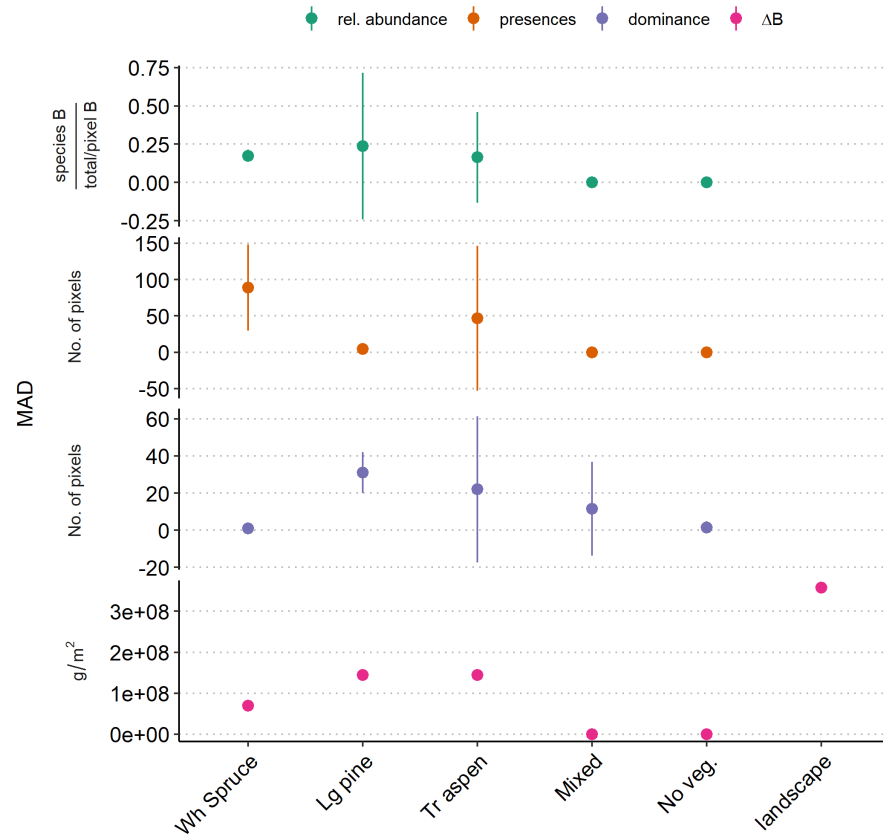


FIGURE 5.2: A plot of landscape-wide mean absolute deviations (MAD) from (top to bottom) observed mean relative abundance, no. of presences, no. of pixels where the species is dominant and ΔB .

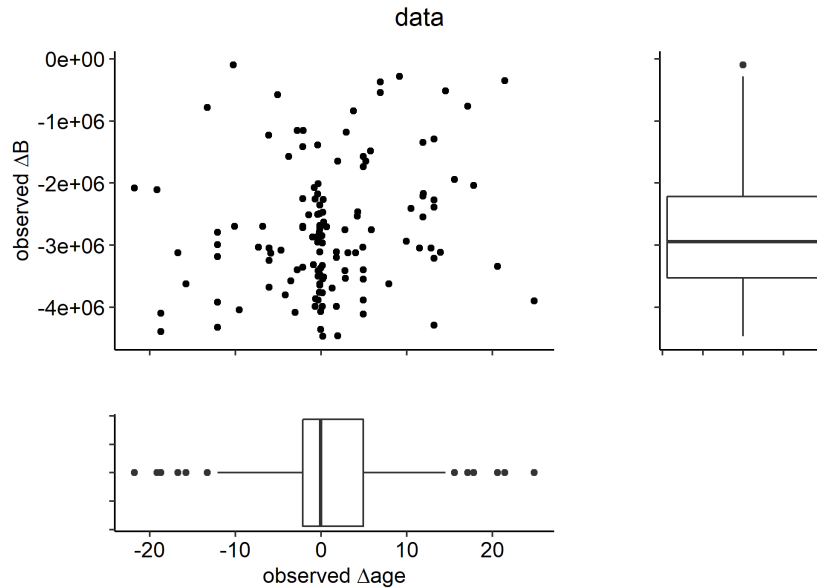


FIGURE 5.3: Diagnostic plot of observed changes in biomass and age ΔB and ΔAge , respectively).

5.4 References

- [18] Eliot J. B. McIntire and Alex M. Chubaty. *SpaDES.experiment: Simulation Experiments Within The SpaDES Ecosystem*. 2021. URL: <https://github.com/PredictiveEcology/SpaDES.experiment>.
- [28] Shelley Stall et al. “Make scientific data FAIR”. In: *Nature* 570.7759 (June 2019), pp. 27–29. DOI: 10.1038/d41586-019-01720-7¹⁹. URL: <http://www.nature.com/articles/d41586-019-01720-7>.
- [29] Brian R. Sturtevant et al. *LANDIS-II Dynamic Fire System Extension v3.0 – User Guide*. 2018.

¹⁹<https://doi.org/10.1038/d41586-019-01720-7>



Bibliography

- [1] Ceres Barros et al. “Empowering ecological modellers with a PERFICT workflow: Seamlessly linking data, parameterisation, prediction, validation and visualisation”. In: *Methods in Ecology and Evolution* 14 (2023), pp. 173–188. DOI: 10.1111/2041-210X.14034²⁰.
- [2] Ceres Barros et al. “Empowering ecological modellers with a PERFICT workflow: seamlessly linking data, parameterisation, prediction, validation and visualisation”. In: *Methods in Ecology and Evolution* 14.1 (2023), pp. 173–188. DOI: 10.1111/2041-210X.14034²¹.
- [3] A Beaudoin et al. “Mapping attributes of Canada’s forests at moderate resolution through kNN and MODIS imagery”. In: *Canadian Journal of Forest Research* 44 (2014), pp. 521–532. DOI: 10.1139/cjfr-2013-0401²².
- [4] A Beaudoin et al. *Species composition, forest properties and land cover types across Canada’s forests at 250m resolution for 2001 and 2011*. 2017. DOI: 10.23687/EC9E2659-1C29-4DDB-87A2-6ACED147A990²³. URL: <http://open.canada.ca/data/en/dataset/ec9e2659-1c29-4ddb-87a2-6aced147a990>.
- [5] P J Burton and S G Cumming. “Potential effects of climatic change on some western Canadian forests, based on phenological enhancements to a patch model of forest succession”. In: *Water, Air and Soil Pollution* 82 (1995), pp. 401–414.
- [6] P J Burton and S G Cumming. “Potential effects of climatic change on some western Canadian forests, based on phenological enhancements to a patch model of forest succession”. In: *Water, Air and Soil Pollution* 82 (1995), pp. 401–414.

²⁰<https://doi.org/10.1111/2041-210X.14034>

²¹<https://doi.org/10.1111/2041-210X.14034>

²²<https://doi.org/10.1139/cjfr-2013-0401>

²³<https://doi.org/10.23687/EC9E2659-1C29-4DDB-87A2-6ACED147A990>

- [7] Alex M. Chubaty and Eliot J. B. McIntire. *SpaDES: Develop and Run Spatially Explicit Discrete Event Simulation Models*. 2019. URL: <https://CRAN.R-project.org/package=SpaDES>.
- [8] Dean W Coble and Young-Jin Lee. "Use of a generalized sigmoid growth function to predict site index for unmanaged loblolly and slash pine plantations in East Texas". In: *Proceedings of the 13th Biennial Southern Silvicultural Research Conference*. Ed. by Kristina F. Connor. Gen. Tech. Rep. SRS-92. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station, 2006, pp. 291–295. URL: <https://research.fs.usda.gov/treearch/23398>.
- [9] John Cosco. *COMMON ATTRIBUTE SCHEMA (CAS) FOR FOREST INVENTORIES ACROSS CANADA*. Feb. 2011, p. 117.
- [10] Desta Fekedulegn, Mairitin P. Mac Siurtain, and Jim J. Colbert. "Parameter estimation of nonlinear growth models in forestry". In: *Silva Fennica* 33.4 (1999). Article id 653. DOI: 10.14214/sf.653²⁴.
- [11] L Guindon et al. *SCANFI: the Spatialized Canadian National Forest Inventory data product*. 2023. DOI: 10.23687/18E6A919-53FD-41CE-B4E2-44A9707C52DC²⁵. URL: <https://open.canada.ca/data/en/dataset/18e6a919-53fd-41ce-b4e2-44a9707c52dc> (visited on 09/26/2025).
- [12] Luc Guindon et al. "A new approach for spatializing the Canadian National Forest Inventory (SCANFI) using Landsat dense time series". en. In: *Canadian Journal of Forest Research* 54.7 (July 2024), pp. 793–815. ISSN: 0045-5067, 1208-6037. DOI: 10.1139/cjfr-2023-0118²⁶. URL: <https://cdnsiencepub.com/doi/10.1139/cjfr-2023-0118> (visited on 09/26/2025).
- [13] Txomin Hermosilla et al. "Characterizing long-term tree species dynamics in Canada's forested ecosystems using annual time series remote sensing data". en. In: *Forest Ecology and Management* 572 (Nov. 2024), p. 122313. ISSN: 03781127. DOI: 10.1016/j.foreco.2024.122313²⁷. URL: <https://linkinghub.elsevier.com/retrieve/pii/S037811272400625X> (visited on 09/26/2025).

²⁴<https://doi.org/10.14214/sf.653>

²⁵<https://doi.org/10.23687/18E6A919-53FD-41CE-B4E2-44A9707C52DC>

²⁶<https://doi.org/10.1139/cjfr-2023-0118>

²⁷<https://doi.org/10.1016/j.foreco.2024.122313>

- [14] Txomin Hermosilla et al. “Land cover classification in an era of big and open data: Optimizing localized implementation and training data selection to improve mapping outcomes”. en. In: *Remote Sensing of Environment* 268 (Jan. 2022), p. 112780. ISSN: 00344257. DOI: 10.1016/j.rse.2021.112780²⁸. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0034425721005009> (visited on 10/02/2025).
- [15] E. H. (Ted) Hogg, J. P. Brandt, and M. Michaelian. “Impacts of a regional drought on the productivity, dieback, and biomass of western Canadian aspen forests”. In: *Canadian Journal of Forest Research* 38.6 (June 2008), pp. 1373–1384. ISSN: 0045-5067, 1208-6037. DOI: 10.1139/X08-001²⁹. URL: <http://www.nrcresearchpress.com/doi/10.1139/X08-001>.
- [16] M.-C. Lambert, C.-H. Ung, and F. Raulier. “Canadian national tree aboveground biomass equations”. In: *Canadian Journal of Forest Research* 35.8 (2005), pp. 1996–2018. DOI: 10.1139/x05-112³⁰.
- [17] Eliot J. B. McIntire and Alex M. Chubaty. *reproducible: A Set of Tools that Enhance Reproducibility Beyond Package Management*. 2020. URL: <https://reproducible.predictiveecology.org,%20https://github.com/PredictiveEcology/reproducible>.
- [18] Eliot J. B. McIntire and Alex M. Chubaty. *SpaDES.experiment: Simulation Experiments Within The SpaDES Ecosystem*. 2021. URL: <https://github.com/PredictiveEcology/SpaDES.experiment>.
- [19] Christian Messier, Sylvain Parent, and Yves Bergeron. “Effects of overstory and understory vegetation on the understory light environment in mixed boreal forests”. In: *Journal of Vegetation Science* 9.4 (Aug. 1998), pp. 511–520. ISSN: 11009233. DOI: 10.2307/3237266³¹. URL: <http://doi.wiley.com/10.2307/3237266>.
- [20] NRCan. *National Ecological Framework for Canada - Terrestrial Ecozones*. 2013.
- [21] Changhui Peng et al. “A drought-induced pervasive increase in tree mortality across Canada’s boreal forests”. In: *Nature Climate Change* 1.9 (Dec. 2011), pp. 467–471. ISSN: 1758-678X, 1758-6798. DOI: 10.1038/nclimate1293³². URL: <http://www.nature.com/articles/nclimate1293>.

²⁸<https://doi.org/10.1016/j.rse.2021.112780>

²⁹<https://doi.org/10.1139/X08-001>

³⁰<https://doi.org/10.1139/x05-112>

³¹<https://doi.org/10.2307/3237266>

³²<https://doi.org/10.1038/nclimate1293>

- [22] Paul D Pickell and Nicholas C Coops. *Development of historical forest attribute layers using Landsat time series and kNN imputation for the western Canadian boreal forest*. Tech. rep. University of British Columbia, Dec. 2016, pp. 1–27.
- [23] Robert M. Scheller and James B. Domingo. *LANDIS-II Model v6.0 – Conceptual Description*. Apr. 20, 2012.
- [24] Robert M. Scheller and James B. Domingo. *LANDIS-II Model v6.0 – User Guide*. July 19, 2011.
- [25] Robert M. Scheller and Brian R. Miranda. *LANDIS-II Biomass Succession v3.2 Extension – User Guide*. 2015.
- [26] Robert M. Scheller and Brian R. Miranda. *LANDIS-II Biomass Succession v3.2 Extension – User Guide*. 2015.
- [27] Robert M. Scheller and David J. Mladenoff. “A forest growth and biomass module for a landscape simulation model, LANDIS: design, validation, and application”. In: *Ecological Modelling* 180.1 (Dec. 2004), pp. 211–229. DOI: 10.1016/j.ecolmodel.2004.01.022³³. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0304380004003837>.
- [28] Shelley Stall et al. “Make scientific data FAIR”. In: *Nature* 570.7759 (June 2019), pp. 27–29. DOI: 10.1038/d41586-019-01720-7³⁴. URL: <http://www.nature.com/articles/d41586-019-01720-7>.
- [29] Brian R. Sturtevant et al. *LANDIS-II Dynamic Fire System Extension v3.0 – User Guide*. 2018.
- [30] Chhun-Huor Ung, Pierre Bernier, and Xiao-Jing Guo. “Canadian national biomass equations: new parameter estimates that include British Columbia data”. In: *Canadian Journal of Forest Research* 38.5 (2008), pp. 1123–1132. DOI: 10.1139/X07-224³⁵.
- [31] Hadley Wickham. “testthat: Get Started with Testing”. In: *The R Journal* 3.1 (2011), p. 5. DOI: 10.32614/RJ-2011-002³⁶. URL: <https://journal.r-project.org/archive/2011/RJ-2011-002/index.html>.

³³<https://doi.org/10.1016/j.ecolmodel.2004.01.022>

³⁴<https://doi.org/10.1038/d41586-019-01720-7>

³⁵<https://doi.org/10.1139/X07-224>

³⁶<https://doi.org/10.32614/RJ-2011-002>

- [32] Mark D. Wilkinson et al. “The FAIR Guiding Principles for scientific data management and stewardship”. In: *Scientific Data* 3.1 (Dec. 2016), p. 160018. ISSN: 2052-4463. DOI: 10.1038/sdata.2016.18³⁷. URL: <http://www.nature.com/articles/sdata201618>.
- [33] Jian Yang et al. “An innovative computer design for modeling forest landscape change in very large spatial extents with fine resolutions”. In: *Ecological Modelling* 222.15 (Aug. 10, 2011), pp. 2623–2630. DOI: 10.1016/j.ecolmodel.2011.04.032³⁸. URL: <https://www.sciencedirect.com/science/article/pii/S0304380011002651>.

³⁷<https://doi.org/10.1038/sdata.2016.18>

³⁸<https://doi.org/10.1016/j.ecolmodel.2011.04.032>