Updating the MYRLIN models for growth projection in mixed tropical forest

Prepared under contract for FAO, Rome

by
Denis Alder
Consultant in Forest Biometrics
March 2020

SUMMARY

This report describes re-analysis of datasets from Brazil, Guyana, Costa Rica and Papua New Guinea that comprised 128,029 tree measurements on 2129 taxa (species and genera) to produce updated models based on the MYRLIN (myrlin.org) system. 15 growth model groups are formed using a weighted k-means analysis, and characterised with the prefixes P, L, M and S, representing pioneer, light-demanding, intermediate, and shade bearing species in four size categories, indicated by suffixes 1-4, being trees with typical mature sizes approximately in the ranges less than 30 cm, 30-60 cm, 60-80-70 cm and 80 cm + dbh. Mature size is indicated by the 95% quantile of the species diameter distribution (D95) on all plots. Additionally, mean increment of the model groups is related to wood density from an international wood density database, and a high correlation (R^2 69%) found. From this a decision-table is developed, based on species size, ecological characteristics and wood properties, to assign species to growth model groups when growth rate information is not directly available from permanent sample plots (PSPs).

All analyses are carried out using the R statistical package, and each workflow script is documented and listed in the report, including the code used to generate all the graphics included. There are figures showing the ordination of the species data by D95 and mean increment (Dinc), the clusters derived from k-means analysis, survival curves, and the correlation and allometric model between increment and wood density.

Mortality rates are also estimated from the data set using Kaplan-Meier survival curves and the R *survival* package. This is very suitable for data in which trees may start and stop being measured at arbitrary points over the study period (as ingrowth, mortality and harvesting occur), and also the entire data set combines PSPs with different measurement periods. Annual mortality rate (AMR) is also tested against indicators or damage and tree dominance, but no clear patterns are found, possibly because of the diversity of methodologies involved in this aggregation of plots.

Both the original consolidated dataset, a cleaned species list of 1985 taxa with assigned growth models, and tables of mean increment, D95, AMR and wood density by growth model group are provided as outputs, downloadable as documented in the report.

The study will contribute to the further updating and elaboration by FAO of MYRLIN as a forest planning tool for mixed tropical forests in situations where locally applicable data is limited.

Contents

Acknowledgements

The work described in this report is based on the long-term field work and meticulous maintenance and measurement of Permanent Sample Plots in Brazil, Guyana, Costa Rica and Papua New Guinea. Although the data used in this analysis has already been described and published in journal articles referenced in the report, the author would like to thank in particular those most closely associated with this effort and with whom he has worked in the field. These include in Brazil, José Natalino M. Silva, João Olegário Pereira de Carvalho, José do Carmo Lopes, Lia C. de Oliveira and Ademir R. Ruschel; in Guyana, Peter van der Hout, Roderick Zagt, Hans ter Steeg and staff of the Guyana Forestry Commission, especially James Singh, Jagdesh Singh and Mohammad Khan; in Costa Rica, Mauricio Sanchez and Stewart Maginnis; and in Papua New Guinea, Cossey Yosi, Forova Oavika and Julian Fox. In FAO, the author has worked closely with Lauri Vesa, who has guided this work, under the departmental direction of Julian Fox, and with administrative support from Esther Phillips.

Disclaimer

This report presents the author's research, analysis and findings as an independent consultant. Any errors or omissions are solely the responsibility of the author. It does not represent any formal position or opinion of FAO or any staff member of the organisation in any respect and should not be construed or referenced as such.

Abbreviations

AMR	Annual Mortality Rate
BR	Brazil
CAFOGROM	Cohort Amazon Forest Growth Model
CR	Costa Rica
CSV	Comma Separated Value (file format)
DBF	Database Format (file format)
dbh	Diameter at breast height (1.3 m)
GY	Guyana
KM	Kaplan-Meier survival analysis, curves
MYRLIN	Method of Yield Regulation with Limited Information
NA	Not available (R data value)
PG	Papua New Guinea
PSP	Permanent Sample Plot
R	R statistical and programming language
SQL	System Query Language (database language)

1. Introduction

This report covers work done by the author under a consultancy contract for FAO from October to December 2019. Its objective was to update to the MYRLIN growth models for data acquired and available since the original publication in 2002, as part of ongoing work in FAO to convert MYRLIN to an R-based system accessible over the internet using the R-Shiny platform. This is likely now to be integrated within the Arena framework for forest management tools that FAO is developing (Vesa, 2019).

MYRLIN is an acronym for Methods of Yield Regulation with Limited Information. The original website and information can be seen via the URL myrlin.org, which is in fact archived on the consultant's website at http://www.denisalder.net/myrlin/original/. The related scientific publications are Alder et al. (2002) and Wright & Alder (2000). MYRLIN consists of 3 Excel workbooks which are free to distribute and modify from the myrlin.org website. The modules cover (1) A forest inventory tool, that can compile stand tables from a variety of inventory designs for mixed tropical forest, (2) a tool to help estimate growth rates given information on species typical size at maturity and species ecology (pioneer, light demanding, or shade bearing), and (3) a concession modelling tool, using output tables generated by modules 1 and 2, that calculates future yields as tables and charts from a forest being managed as a series of coupes over a felling cycle. The modelling tool is based on simple stand projection using diameter classes and constant growth rates by species groups.

That data sets used in the original work are described in Alder et al (2002), and come from PSP dispositions in Eastern Amazonia (Tapajos and Jari), in Guyana (Pibiri), from several contributors in Costa Rica, and from PSPs in Papua New Guinea managed by the Forest Research Institute at Lae. Three of these data sets have been remeasured several times, and the objective of this report was to update and revise the growth models, and provide better tools for estimating growth rates based on this information.

To this end, the report describes the consolidation of the updated databases from various sources, the cleaning and updating of species names, the re-calculation of growth model groups with growth rates and mortality factors for each group and the relationship between these and wood density. It provides a number of data sets and tables as outputs that can be used to estimate growth rates, given information on species size, ecology and wood density. These can all be downloaded as either R data files or as CSV files.

All analysis has been done using the R statistical language, to provide compatibility with FAO's current methodology for forest growth and yield work. Each stage of analysis has been packaged as a workflow, or series of steps, contained within a script. Five scripts are documented and listed in this report, and by following them, with the original datasets, the original analysis and all the figures shown can be reconstructed. Apart from documenting the work and methods, this may also provide a useful didactic tool for training workshops in R and data analysis.

The original datasets of forest growth by species and plots contain a considerable body of work of scientific potential. Some of this has been published in Alder et al (2012) and Fox et al (2011), but in order to protect the interests of the data owners relative to other potential scientific outputs, many details of the plot history, experimental design and precise locations are not included in the data file, nor are intermediate measurements or detailed coding and scoring of a number of factors. Only the minimal and simplified data that formed part of the MYRLIN study are included.

2. Preparation of consolidated datasets

2.1. Description of datasets included

For this study, the original data sets used for the MYRLIN model in 2002 (Alder et al, 2002, Table 1) have been extended with new measurements. These data sets are from permanent sample plots (PSPs) in Eastern Amazonia (Brazil), Pibiri in central Guyana, Costa Rica, and Papua New Guinea. The new data includes measurements mainly up to 2007-8, and for the Guyana plots, up to 2013. All data are from lowland humid mixed tropical forest, and all data are from plots which have undergone some degree of exploitation, either through commercial logging or as experimental treatments. The numbers, sizes, and measurement thresholds of the various plots are shown in Table 1 below. Where a plot has sub-plots for measuring smaller diameter trees, the sub-plot size and minimum diameter is shown in square parentheses.

Table 1 : Permanent Sample Plot locations included in the consolidated data set

Code	Location	Plot Size (ha)	Minimum DBH (cm)	No. PSPs	No. Trees	Trees 20 cm+ dbh	No. Genera	Years Measured
Brazil – Eastern Amazonia (BR)								
3	Tapajós Forest, Pará, Brazil, km 67 (RP 12). Primary forest, conventional logging in 1979.	0.25	5	36	17232	2335	164	1981-2007
4	Tapajós Forest, Pará, Brazil, km 114 (RP 11), Primary forest, controlled exploitation with various logging and thinning treatments.	0.25	5	48	25876	3636	173	1981-2008
5	Jarí, Amapá, Brazil (RP 21). Primary forest, various controlled felling treatments in 1985.	1	20	36	7737	7651	163	1984-1996
6	Jarí, Amapá, Brazil (RP 22). Secondary forest succession. Experiment, clearfelled 1982, some treatments applied.	1 [0.1]	20 [5]	8	7245	233	109	1985-1996
11	Jarí, Amapá, Brazil (RP 25). Primary forest, unlogged control plots.	1 [0.1]	20 [5]	4	905	787	106	1984-1996
12	Jarí, Amapá, Brazil (RP 26). Secondary forest succession. Clear-felled 1982, no further treatment. Control to RP 22.	1 [0.1]	20 [5]	4	1505	236	122	1985-1996
Costa Rica (CR)								
1	CODEFORSA 1-ha plots	1	20	9	4464	1427	227	1991-1998
2	CODEFORSA ¼-ha plots	0.25	20	27	3724	987	160	1992-1998
3	Portico/Tecnoforest plots	1	10	17	7576	2862	129	1990-1996
4	ITCR/DFID Plots	0.49	10	41	9694	3394	169	1994-1997
5	COSAFORMA/GTZ plots	0.06	5	34	2291	276	77	1991-1996
Guyana (GY)								
	Pibiri GFC/Tropenbos plots	1.96 [0.25]	20 [5]	15	11434	7260	117	1993-2013
Papua New Guinea (PG)								
	FRI/ITTO Plots	1	10	72	28346	10487	294	1990-2009
Total				351	128029	41688	683	1981-2013

The Brazil dataset is as used in Alder *et al.* (2012) relative to the updating of the CAFOGROM forest model (Alder & Silva, 2000) . The plot design, layout, history and earlier results are also described in Silva *et al.* (1995, 1996). The Costa Rica datasets are unchanged from the original MYRLIN study (Alder *et al*, 2002, Wright & Alder, 1999). The Guyana datasets are originally described in van der Hout (1999, 2000a, b), with the updated measurements and their features discussed in detail in

Alder (2017). The Papua New Guinea plots and the updated measurements are discussed in Fox *et al* (2011).

For convenience, these datasets will be referred to by their 2-letter country ISO codes, as shown in Table 1, being BR (Brazil), CR (Costa Rica), GY (Guyana) and PG (Papua New Guinea).

2.2. Consolidated Tree Increment Data

These various data sets were mostly originally in FoxPro DBF files, with the updates as Excel tables, and the PG data as Access tables. The BR data had been converted into a MySQL database. The measurement standards and coding of each data set varied, but in all cases, diameter was measured at dbh or above buttress, mortality was recorded, and there was some indication of damage or disease. The various individual datasets were imported into MySQL with their native structures, and then converted via interactive SQL commands into a single common format, with the structure shown in Table 2.

Table 2 : Structure of the consolidated TREEINC data table in MySQL

Field name	Data type	Description
`CC`	`VARCHAR(2)`	*ISO alpha-2 country code for data origin*
`PLOTID`	`VARCHAR(7)`	*Plot ID, unique within a country. The original plot IDs in the parent datasets were converted where necessary for uniqueness. The prefix number corresponds to the numbers in Table 1, column 1.*
`TNO`	`BIGINT(20)`	*Tree number, unique at least with a plot. This may combine tree and quadrat numbers.*
`SPNO`	`INT(11)`	*A unique species number, linking to spno in the consolidated species list (Table 3).*
`SPP`	`VARCHAR(9)`	*The original species code in the species list of the parent dataset. Where these were numeric, they are prefixed with zeroes.*
`DATE1`	`DATE`	*The first measurement date the tree was observed.*
`DATE2`	`DATE`	*The last measurement date the tree was observed. If the tree was not found, this is the date of plot measurement when the tree absence was recorded.*
`DIAM1`	`DOUBLE`	*The diameter (dbh or dab) at first measurement date, in cm.*
`DIAM2`	`DOUBLE`	*The diameter at last measurement date, in cm. This may be null if the tree was not found.*
`INGR`	`TINYINT(4)`	*1 if tree was ingrowth (DATE1> DATE1 for plot), 0= not ingrowth.*
`MORT`	`TINYINT(4)`	*1 if tree died for any reason, 0=alive. Tree harvested or killed by thinnings were not counted in this category.*
`CPOS`	`TINYINT(4)`	*Crown position, 1 for canopy/emergent trees, 0 for mid-canopy or understorey.*
`DMG`	`TINYINT(4)`	*1 if any damage or disease prior to death, 0=healthy, undamaged.*
`ERRFLAG`	`TINYINT(4)`	*1 if diameter data likely unreliable/erroneous. Set when point of measurement changes or increment outliers, or if DATE1=DATE2.*

This table comprised in total 128, 029 records. It was exported to a CSV file called *treeinc.csv*, and then imported into an R dataframe, also called *treeinc*. In R, the character fields convert into factors, and the various number formats into numeric data.

2.3. Consolidated Species List

Each of the regional datasets shown in Table 1 has its own species list. These were combined into a single table, *splist* with region, genus, species and family name, and original species code, as per Table 3. The species codes were in some cases numeric and in others alphabetic, typically based on some abbreviation of genus and species. Numeric codes were right justified and packed with leading zeroes and converted to character codes for compatibility of format with the character-based codings. An arbitrary unique serial number was then applied to all species, called *spno*. Once assigned, it was also written into the same field of the *treeinc* table. The consolidated list contained 2129 entries, from 683 genera. Many of these entries are synonyms or duplications, and as discussed in section 2.4, the list was cleaned up as much as possible before being used for analysis.

Table 3 : Structure of the consolidated species list

Field name	Data type	Description
SPNO	INT(11)	*A unique species number, linking to* SPNO *in the* TREEINC *table (Table 2). This is the primary key and is unique in the table.*
CC	VARCHAR(2)	*ISO alpha-2 country code for data origin*
SPP	VARCHAR(9)	*The original species code in the species list of the parent dataset. Where these were numeric, they are prefixed with zeroes.*
FAMILY	TEXT	*Family name as entered in the original list. This may include various abbreviations.*
GENUS	TEXT	*Genus as given in the original list.*
SPECIES	TEXT	*Species as given in the original list. This frequently involves notes if identification is inexact, such as 'cf. variflora', 'sp. like variflora', and can include sub-species or varieties.*

2.4. Export of data and location of archive files

Once the consolidated database tables TREEINC and SPLIST had been created, they were exported into R dataframes for further processing as CSV files, keeping the same structure as noted above. The files are archived at the URL http://denisalder.net/proj/fao/, hereafter referred to as ~/, and can be downloaded as follows:

~/treeinc.csv .... The tree increment data, as per Table 2
~/splist.csv .... The consolidated species list, per Table 3

The import of the datafiles in R is done with:

```
# import tree data and species list
treeinc <- read.csv("http://denisalder.net/proj/fao/treeinc.csv")
splist <- read.csv("http://denisalder.net/proj/fao/splist.csv")
```

This creates the R dataframes TREEINC and SPLIST with the same column headings as shown in Tables 1 and 2.

2.5. Cleaning the species list

The original species lists contain many out of date and misspelt botanical names. To clean these up an R script provided by Lauri Vesa (Vesa, 2020) was used that checks names against several plant databases. The methodology is also discussed in Perez (2017) in an online presentation. This script has been packaged here as *ValidateSpecies4.r*. The R script shown below shows the various pre- and post processing steps used to run the *ValidateSpecies4.r* script. Lines 3-8 show the various required libraries. In many R sessions it is not explicitly necessary to allocate these, but the lists is shown for completeness. At line 10-18, the splist.csv file is read in, and then a second data frame created, genspp, in which the separate genus and species names are concatenated as scientific_name, and the species code renamed as code, for compatibility with the species checking script. This data is them de-duplicated (any duplicate lines removed), and saved as the file uniqsp.csv, which is the input file expected by the *ValidateSpecies4.r* routine. This is executed at line 20 (the comment prefix # should be removed in operational use – it has been inserted to avoid accidental initiation of this process, which takes 1-2 hours to run.

Script 1 : Workflow for validation and correction of species names

```
 1  # workflow for species checking - 12-15 Feb 2020
 2  # libraries required
 3  library(magrittr)
 4  library(dplyr)
 5  library(data.table)
 6  library(crayon)
 7  library(glue)
 8  library(stringr)
 9  # import latest splist from online repository
10  splist <- read.csv("http://denisalder.net/proj/fao/splist.csv", stringsAsFactors = F)
11  # convert to format for data cleaning
12  genspp <- splist %>% filter(!grepl('unident', genus)) %>%
13    transmute(code=spno, scientific_name=if_else(species=='~', genus, genus %+% ' ' %+% trim(species)))
14  # remove any duplicates
15  dupspp <- duplicated(genspp[,2])
16  uniqsp <- genspp[!dupspp,1:2]
17  # write as CSV (becomes input to ValidateSpecies)
18  write.csv(uniqsp,file='uniqsp.csv', row.names=F)
19  # ------- at this point, run script ValidateSpecies4.R --------
20  # source('T:/FAO/Work/SpeciesCheck/ValidateSpecies4.R')
21  # ------------------------------------------------------------
22  # create df 'cleansp' from sp.accepted with either NewName, if given, or SuggestedName, if given or original name
23  # species code also converted back to integer for compatibility with spno and later usage
24  cleansp <- sp.accepted %>% transmute(code=as.integer(code), spname= if_else(NewName>'',NewName,
25        if_else(suggestedName>'', suggestedName, scientific_name)))
26  # extend genspp to include cc column from original splist. 'code' renamed 'spno'
27  genspp %<>% left_join(splist, by=c('code'='spno')) %>% transmute(code, cc, scientific_name)
28  # add new name from cleansp, rename 'scientific_name' to oldName
29  genspp %<>% left_join(cleansp, by='code') %>% transmute(code, cc, oldName=scientific_name, newName=spname)
30  # de-duplicate rows on 'code'
31  dupcodes <- duplicated(genspp[,1])
32  genspp <- genspp[!dupcodes,]
33  # correct NA in newName by substituting oldName
34  genspp %<>% mutate(newName=if_else(is.na(newName), oldName, newName))
35  # split out genus and species from newName
36  gs <- str_split_fixed(genspp$newName, '\\s+', n=2)
37  genspp %<>% mutate(genus=gs[,1], species=gs[,2])
38  # at this point genspp contains:
39  # - the numeric species code (matching spno in treeinc)
40  # - country of origin for the data (cc)
41  # - name in original species list (oldName)
42  # - cleaned scientific name (newName)
43  # - name split into genus and species
44  # this table is written out as a csv here for archival
45  write.csv(genspp,file='genspp.csv', row.names=F)
46
```

The output from the species validation is in a data frame called *sp.accepted*. At line 24, the essential results are copied back into a dataframe called *cleansp* with the columns *code* and *spname*. *Code* is the original species code (*spno* in *treeinc* and *splist*, tables 2 and 3). The species name is one of the following possibilities from the *sp.accepted* output: (1) *NewName* if given, or *SuggestedName*, if *NewName* is blank, or if both are blank, the original species name is used.

Once this has been done, at lines 27-37, the information is written back to the *genspp* table and various modifications made. The columns are renamed *oldName* (as provided originally), and *newName*, as output from the name validation process. In many cases the two may be the same, especially for the Guyana species, where the botanical names were recently checked and updated. At lines 30-32, duplicates that have arisen from the table joins are removed. At lines 33-34, any species where the new name is NA in R are filled in using the old name. Finally the new binary names are split into separate Genus and Species columns at line 36-37, and the final working table written out as *genspp.csv*. The above workflow script was saved as *Workflow_to_clean_species_names.r*

These various files mentioned above are archived at, and can be downloaded from http://denisalder.net/proj/fao/ as follows, substituting the URL for ~:

~/genspp.csv
~/ValidateSpecies4.r

~/Workflow_to_clean_species_names.r

Note that case is significant in the filenames and should be given as shown.

2.6. Tree increment calculation and consolidation of taxa

The next step in processing was to calculate diameter increment for each tree, set an error flag to indicate which increment observations were unusable, and group species into genera by country. The R script to perform these steps is called workflow_for_species_stats.r and is reproduced in Script 2 below.

Script 2 : Workflow to calculate tree increment and consolidate species as genus/country groups

```
1   # standard libraries used - includes ggplot2, dplyr, stringr etc.
2   library("tidyverse")
3   # magrittr adds %<>% operator
4   library("magrittr")
5   # import tree data and splists
6   treeinc <- read_csv("http://denisalder.net/proj/fao/treeinc.csv")
7   splist <- read_csv("http://denisalder.net/proj/fao/splist.csv")
8   genspp <- read_csv("http://denisalder.net/proj/fao/genspp.csv")
9   # rename code in genspp to spno for consistency and easier joins
10  genspp %<>% rename(spno=code)
11  # calculate dinc column
12  treeinc %<>% mutate(dinc=(diam2-diam1)/as.numeric(date2-date1)*365)
13  # set error flag where dinc not finite (NA, NaN), negative or greater than 5 cm/yr
14  treeinc %<>% mutate(errflag= errflag |if_else(!is.finite(treeinc$dinc) | treeinc$dinc<0 | treeinc$dinc>5, 1, 0))
15  # create table of genus names and country codes with serial number gcno
16  genspp %>% group_by(cc,genus) %>% summarise(ns=n()) %>% arrange(genus, cc) %>% ungroup() %>%
17    mutate(gcno = row_number()) -> gencodes
18  # write the gencode number into a column in genspp and thence treeinc
19  genspp %<>% left_join(select(gencodes, cc, genus, gcno) ,by=c('cc', 'genus'))
20  treeinc%<>% left_join(select(genspp, spno, gcno), by='spno')
21  # make tables of mean increment and D95 for each country/genus (taxon)
22  treeinc %>% filter(errflag==0) %>% group_by(gcno) %>% summarize(dinc=mean(dinc, trim=0.05, na.rm=T)) -> txdinc
23  treeinc %>% filter(errflag==0) %>% group_by(gcno) %>% summarize(d95=quantile(diam1, 0.95, na.rm=T)) -> txd95
24  treeinc %>% filter(errflag==0) %>% group_by(gcno) %>% summarize(nt=n()) -> txnt
25  # combine separate columns into one table, gcstats
26  gencodes %>% left_join(txnt, by='gcno') %>% left_join(txdinc, by='gcno') %>% left_join(txd95, by='gcno') %>%
27      select(gcno, cc, genus, ns, nt, dinc, d95) %>% arrange(desc(nt)) -> gcstats
28  # also get number of trees by species into genspp for reference
29  treeinc %>% group_by(spno) %>% summarise(nt = n()) %>% right_join(genspp, by='spno') %>%
30    select(spno, gcno, cc, oldName, newName, genus, species, nt) %>% arrange(genus, species, cc) -> genspp
```

In this script, lines 1-8 are preliminaries, setting the libraries used and reading in data objects saved at the previous step (see section 2.5). At line 10, the species *code* in the genspp dataframe is renamed as *spno* for compatibility with the original *treeinc* file (it had been changed to code for compatibility with the species validation script of Vesa (2020). Lines 11-12 creates a column in *treeinc* called *dinc* that is the difference between first and last observed diameters, divided by the measurement time interval in years. Lines 13-14 set the error flag variable (*errflag*) according to the following conditions:

- If it is already set TRUE, it remains set. The error flag was set in the MySQL preliminaries if coded notes indicated an unreliable or optically estimated diameter (high buttresses, fluting, climber growth etc), or a change in diameter point of measurement between first (*diam1)* and last (*diam2*) diameter measurements.
- It is set if the diameter increment is negative or exceeds 5 cm per year. Although both cases are possible, they generally indicate bad measurements for long series data as used here.
- It is set true if increment is not computable. This is generally the case if the tree is observed only once, *ie.* it has disappeared after the first measurement.

Lines 15-20 create a table called *gencodes*, which groups species by genus and country code. The reason for this is that frequently in the data species are only identified to genus level, and it is confusing to classify a genus, and a species identification for the same genus, separately in the later species ordination, discussed below. However, in different regions the principal or most common species in a genus will differ, so the country code (cc) is used to maintain them separately in the analysis. The code for the genus/country taxa are called *gcno*, and after being allocated in the *gencodes* table (lines 17-18) are written back into *genspp* (line 20) and thence into *treeinc* (line 21).

In lines 21-24, three temporary tables are created called *txdinc, txd95* and *txnt.* These contain, respectively, the diameter increment, 95% quantile of the diameter distribution, and the count of observed trees, all summarised by *gcno* (genus/country taxon). At lines 25-27 these statistics are merged, together with genus name and country code from the *gencodes* table, to a new tables called *gcstats*. Table 4 shows the first 10 lines of this table, ordered with most common taxon (by tree count, *nt*) first. The column *ns* is the number of species identified in the genus (which may include genus only identification). In total, there are 1008 records in this table.

Table 4 : Listing of first 10 lines of *gcstats* dataframe

```
   gcno cc    genus          ns    nt   dinc   d95
  <int> <chr> <chr>        <int> <int>  <dbl> <dbl>
1   180 BR    Cecropia         4  6171 0.458   12.4
2   488 BR    Inga            13  3444 0.306   24.5
3   957 BR    Unidentified    35  3052 0.155   43.8
4   807 BR    Rinorea          4  2916 0.0671  25.9
5   702 CR    Pentaclethra     1  2637 0.513   66.0
6   760 BR    Protium         16  2148 0.165   27.8
7   354 BR    Eschweilera     10  1727 0.113   41.8
8   859 BR    Sloanea          4  1510 0.162   20.1
9   631 PG    Myristica       10  1125 0.261   27.4
10  902 BR    Tachigali        8  1091 0.404   55.1
```

Lines 28-30 are for later reference in reporting some results by species, rather than by genus/country, and calculate the number of trees for each species, putting the result in the *nt* column of the *genspp* table.

3. Forming Growth Model Groups

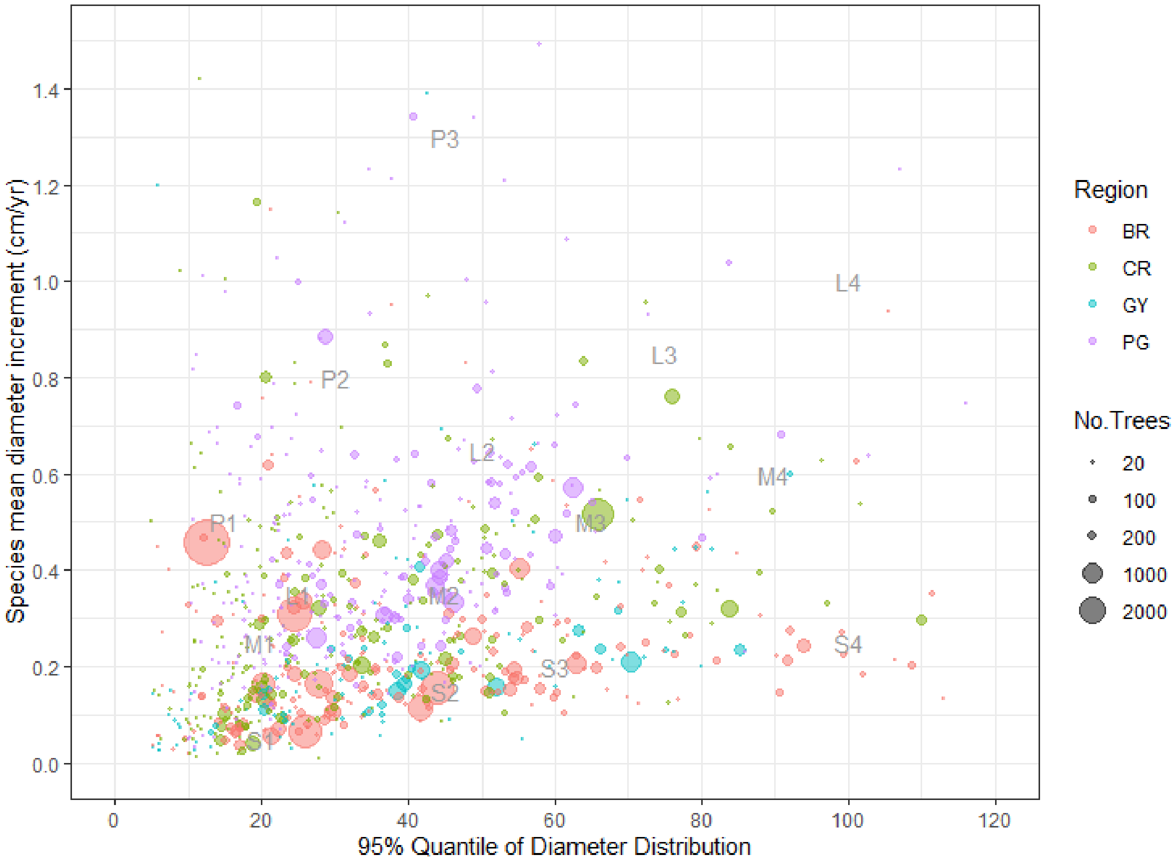
3.1. Ordination method and choice of initial centroids

The method of forming growth model groups is that described in Alder et al. (2002). It is a k-means ordination analysis with pre-set starting centroids and with groups weighted by the number of trees in each group. The ordination is by 95% quantile of the diameter distribution (x), as a robust statistic of the representative size that mature individuals of the species may attain, versus the mean diameter (y). Script 3 shows the R workflow involved.

The data summarised by genus/species codes is in the table *gcstats*. Figure 1 shows a plot of this data, with circle are being proportional to the number of sample trees. Each point is one taxon. Colours show the region of origin. Lightly superimposed in feint are the initial centroid positions, estimated by eye, for the growth model groups.

Figure 1 shows the data with initial centroids, estimated by eye. Each point represents one taxon (genus+country), with the size of the point proportional to the number of trees in that taxon. The R code to produce Figure 1 is at lines 17-24 of Script 3. The centroids (growth model groups) are designated by prefix letters P, L, M, S, indicating predominantly pioneer, light-demanding, median shade tolerance, and shade bearing species. The suffixes 1-4 relate to typical size (95% quantile of the diameter distribution, D95), from small (1, around 20 cm dbh) to 4 (large, around 90-100 cm dbh).

Figure 1 Diameter increment-species size ordination data with initial centroid estimates for species groups



Script 3 : Workflow to form growth model groups and related graphs

```
# workflow covering plotting of charts of dinc on d95, centroid and cluster fitting,
# graphs of cluster data and text graphs of groups
#---------------
# clear existing workspace and load data objects from previous step
rm(list=ls())
load("T:/FAO/Work/Species Models/workflow_for_species_stats.rdata")
# standard libraries used - includes ggplot2, dplyr, stringr etc.
library("magrittr")
library("tidyverse")
# tidy uop unwanted tx* objects
rm(list=ls(pattern='tx.+'))
# read in list of centroid starting values
# centroids2.csv - initial list created from visual inspection of data
centroids = read_csv('centroids2.csv')
centroids %<>% rename(gm=group)
# make dinc/d95 diagram for genus/country taxa
fig1 <- gcstats  %>% filter(dinc>0, d95>0, nt>0) %>%
   ggplot() + geom_point(aes(x=d95, y=dinc, size=nt, color=cc), alpha=0.5) +
   geom_text(data=centroids, mapping=aes(x=d95, y=dinc, label=gm), size=4, color="Grey60") +
   labs(x="95% Quantile of Diameter Distribution", y="Species mean diameter increment (cm/yr)", color="Region") +
   scale_size(breaks=c(20, 100, 200, 1000, 2000), range=c(0.1,10), name="No.Trees") +
   scale_x_continuous(limits=c(0,120), breaks=c(0, 20,40,60,80, 100, 120)) +
   scale_y_continuous(limits=c(0,1.5), breaks=c(0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4)) +
   theme_bw()
# display plot
fig1
# --- steps for allocating taxa to centroids and adjusting them ----
library("raster")       # fast distanceToPoint function
options(warn=-1)        # turn off warning messages (too many within loops)
# create filtered version of gcstats excluding missing data
gcstats  %>% filter(dinc>0, d95>0, nt>0) -> gcstatsf
```

```
32  # Dinc/D95 data for the genera are re-normalized and placed in sdata
33  gcstatsf  %>% transmute(gcno, rx=scale(d95), ry=scale(dinc), cid=NA) -> sdata
34  # scaling factors for d95 and dinc to apply to centroids
35  gcstatsf  %>% summarize(mean=mean(d95), sd = sd(d95)) -> sfx
36  gcstatsf  %>% summarize(mean=mean(dinc), sd = sd(dinc)) -> sfy
37  # scale centroids by these factors, result in scntr
38  centroids %>% transmute(cid = row_number(), rx=scale(d95, center=sfx[1], scale=sfx[2]),
39                      ry=scale(dinc, center=sfy[1], scale=sfy[2])) -> scntr
40  # next operations are repeated until change in scaled centroids (delta) between iterations is less than 0.001
41  delta <- 1
42  iter <- 0
43  gcstats$cid <- NA
44  while(delta>0.001 & iter<20){
45     # calculate all possible distances between between scaled species data and centroids
46     dmx <- pointDistance(sdata[,2:3], scntr[,2:3], lonlat = F)
47     # get index of closest centroid and write into 'cid' column of sdata
48     for(i in 1:nrow(dmx)){
49        # filter out any rows which contain NA
50        if(!any(is.na(dmx[i,]))){
51           # get index of minimum distances and write it into 'cid' column for species data
52           k <- which.min(dmx[i,])
53           sdata[i,'cid'] <- k
54        }
55     }
56     # copy centroid ID into gcstats
57     gcstats %<>% left_join(sdata, by='gcno') %>% transmute(gcno, cid=cid.y, cc, genus, ns, nt, dinc, d95)
58     # calculate weighted sums of new centroids in table inc.models
59     gcstats %>% filter(dinc>0, d95>0, nt>0) %>% group_by(cid) %>% summarise(d95=sum(d95*nt,na.rm=T)/sum(nt),
60                                dinc=sum(dinc*nt,na.rm=T)/sum(nt), nt=sum(nt), ng=n()) -> inc.models
61     # get relative change in scaled centroids from last iteration
62     scntr0 <- scntr
63     scntr <- inc.models %>% transmute(cid, rx=as.vector(scale(d95, center=sfx[1], scale=sfx[2])),
64                                    ry=as.vector(scale(dinc, center=sfy[1], scale=sfy[2])))
65     # sum of Euclidean distance between scaled centroids
66     delta <- sum(sqrt((scntr0[,'rx']-scntr[,'rx'])^2+(scntr0[,'ry']-scntr[,'ry'])^2))
67     # display iteration progress
68     iter <- iter + 1
69     cat("Iteration", iter,"   Delta", delta,"\n")
70  }
71  # add model text using cid index to gcstats and inc.models
72  inc.models$gm <- centroids$gm
73  gcstats %<>% left_join(inc.models, by='cid') %>% transmute(gcno, gm, cid, cc, genus, ns, nt=nt.x, dinc=dinc.x,
74                       d95=d95.x)
75  # outputs
76  cat("inc.models contains final list of centroids. nt is no. sample trees, ng is no. of genera\n")
77  print(inc.models)
78  # graph with identical scaling etc to fig1 showing fitted centroids and coloured by model group
79  fig2 <- gcstats  %>% filter(dinc>0, d95>0, nt>0) %>%
80     ggplot() + geom_point(aes(x=d95, y=dinc, size=nt, color=gm), alpha=0.5) +
81     geom_text(data=inc.models, mapping=aes(x=d95, y=dinc, label=gm), size=4, color="Black") +
82     labs(x="95% Quantile of Diameter Distribution", y="Species mean diameter increment (cm/yr)", color="Model") +
83     scale_size(breaks=c(20, 100, 200, 1000, 2000), range=c(0.1,10), name="No.Trees") +
84     scale_x_continuous(limits=c(0,120), breaks=c(0, 20,40,60,80, 100, 120)) +
85     scale_y_continuous(limits=c(0,1.5), breaks=c(0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4)) +
86     theme_bw()
87  fig2
```

In Script 3, lines 27-40 set up the initial conditions for adjusting the initial estimates of group centroids for the data. The analysis has to be done in a scaled, normalized data space so that the scale differences between diameter (x coordinate) and increment (y coordinate) do not influence the outcome. A table *sdata* is created at line 33 which has the taxon index (*gcno*) and the normalized equivalents of diameter increment and D95 as *rx* and *ry*, respectively. The scaling parameters used, being means and standard deviations of Dinc and D95, are saved as vectors *sfx* and *sfy* in lines 35-36, and then applied to give scaled centroid coordinates at line 38-39 in a table *scntr*.

At lines 41-43, the value *delta*, which represents the change in total sums of squares of differences between the current and last estimates of the centroid positions, is initialised to an arbitrary large number. The iterations end when this is less than 0.001 (line 44). Iterations are also counted by variable *iter*, and the process will stop after at most 20 iterations. At line 43 an empty column of group indices *cid* is added to the table *gcstats*, to simplify later coding.

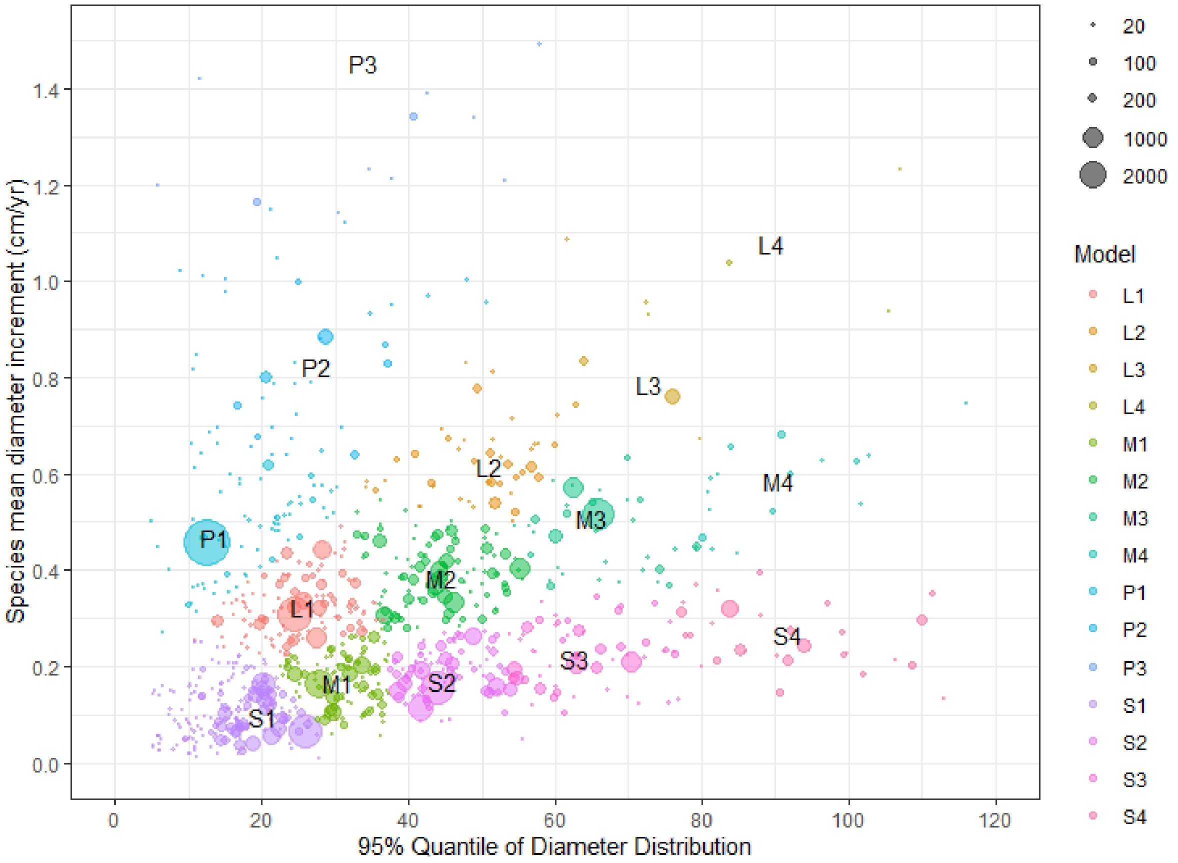
The process then iterates, from lines 44 to 70, until either the limit of iterations is reached, or the convergence criterion is satisfied. At line 46 a matrix *dmx* is calculated whose rows are taxon IDs and columns are centroid IDs. It contains the Euclidean distance in the scaled ordination space between every taxon and every centroid. This is computed via the function *pointDisance* from the library *raster*. Note that the *lonlat* parameter must be set False to give Euclidean distances, otherwise geoid geometry would be used.

Lines 48-54 find, for every row in dmx (taxon), the column (centroid) that has the least distance, and put its index in the table *sdata* for that taxon. Once this has been done for all taxa, then at lines 56-60 the index of the closest centroid is copied back into the table *gcstats*, and the centroids re-computed as the mean D95 and Dinc for all the taxa in each group. At lines 61-66 the shift in distances between centroids from the current and previous iterations is computed as *delta*. Lines 67 display progress and count iterations.

Once this iterative process has terminated, at line 71, then the text IDs of the centroids ('L1', 'L2' etc) are copied back into the table *inc.models*, which holds the final estimates for the recalculated group centroids. At line 73 these text IDs are also copied back into the table of taxon data, *gcstats*. Finally the fitted group centroids are displayed at line 77. This is shown here in Table 5 below. Lines 79-87 plot the results, as shown in Figure 2.

Figure 2 Taxon data with fitted growth model groups

Taxa are Genus/Region combinations, and may include several species.
Model are shown by colour, with a centroid under the model letters shown.



3.2. Results of fitting species groups

Figure 2 shows the genus/country taxa as points, coloured according to growth model group, with the fitted centroid under the letter codes. Table 2 below shows the data for all model groups.

Table 2 : Growth model groups mean diameter increment and 95% diameter quantile

Model Code *gm*	Model Index *cid*	95% diameter quantile, cm *d95*	Mean diam. growth, cm/yr *dinc*	Number of trees *nt*	Number of genera *ng*
S1	1	20.1	0.097	16,449	187
S2	2	44.6	0.171	14,510	78
S3	3	62.7	0.217	6,952	54
S4	4	91.7	0.268	3,552	33
M1	5	30.4	0.165	10,134	101
M2	6	44.6	0.384	13,706	98
M3	7	65.1	0.508	5,664	34
M4	8	90.6	0.586	640	17
L1	9	25.8	0.324	12,465	113
L2	10	51.1	0.615	2,765	42
L3	11	72.7	0.787	880	7
L4	12	89.4	1.080	61	3
P1	13	13.6	0.467	7,305	67
P2	14	27.5	0.822	1,638	37
P3	15	33.9	1.450	434	16

As noted above, these groups are designed to have an obvious interpretation. The S prefix are all slower growing, with diameter increments ranging from 1-3 mm/yr, and in order to survive in the mixed forest environment, will necessarily be shade tolerant, whilst the slow growth will be associated with high wood density. However, the larger species in the S4 group will typically be very long-lived emergents, perhaps several hundred years old at their mature size.

Conversely, the L group will be light-demanding, faster growing (3-10 mm/yr) species, with less dense wood. The largest-sized species will be typically gap opportunists, able to persist in the upper canopy or as emergents, whilst small ones will be associated with gaps and disturbed sites. The P group are the true pioneers, relatively ephemeral in forest terms (usually disappearing after 20-40 years), unable to persist in shade, and occurring on highly disturbed sites.

3.3. File downloads

The files listed below are available to download from the above analysis:

- All data tables and intermediate variables described in this section, including plots for Figs 1 and 2, as R objects:

  http://denisalder.net/proj/fao/workflow_for_growth_model_groups_and_figures.rdata

- Code for Script 3 to estimate growth model centroids.

  http://denisalder.net/proj/fao/workflow_for_growth_model_groups_and_figures.r

4. Survival and Mortality Analysis

4.1. Workflow for Kaplan-Meier Survival functions and mortality rates by species groups

The *treeinc* table (see Table 2) has indicators for mortality, being 1 for trees which have died during the measurement period, and 0 for those that remain alive. This can be considered as a typical right-censored survival study, with individuals (trees) being enrolled at the start of the study (PSP establishment) or recruited at some point after its initiation (ingrowth to minimum diameter, recruitment), and either surviving to the end point of the study or dying during the study period. This is well suited to Kaplan-Meier survival analysis (Wikipedia, 2020; Kaplan & Meyer, 1958), which is now quite frequently used for forestry mortality studies and is well supported in the R Survival package.

Script 4 below shows the workflow for computing the survival curves for each growth model curve, estimating from these the average annual mortality rates (AMR) for each group, and plotting both the survival and mortality curves, as shown in Figures 3 and 4.

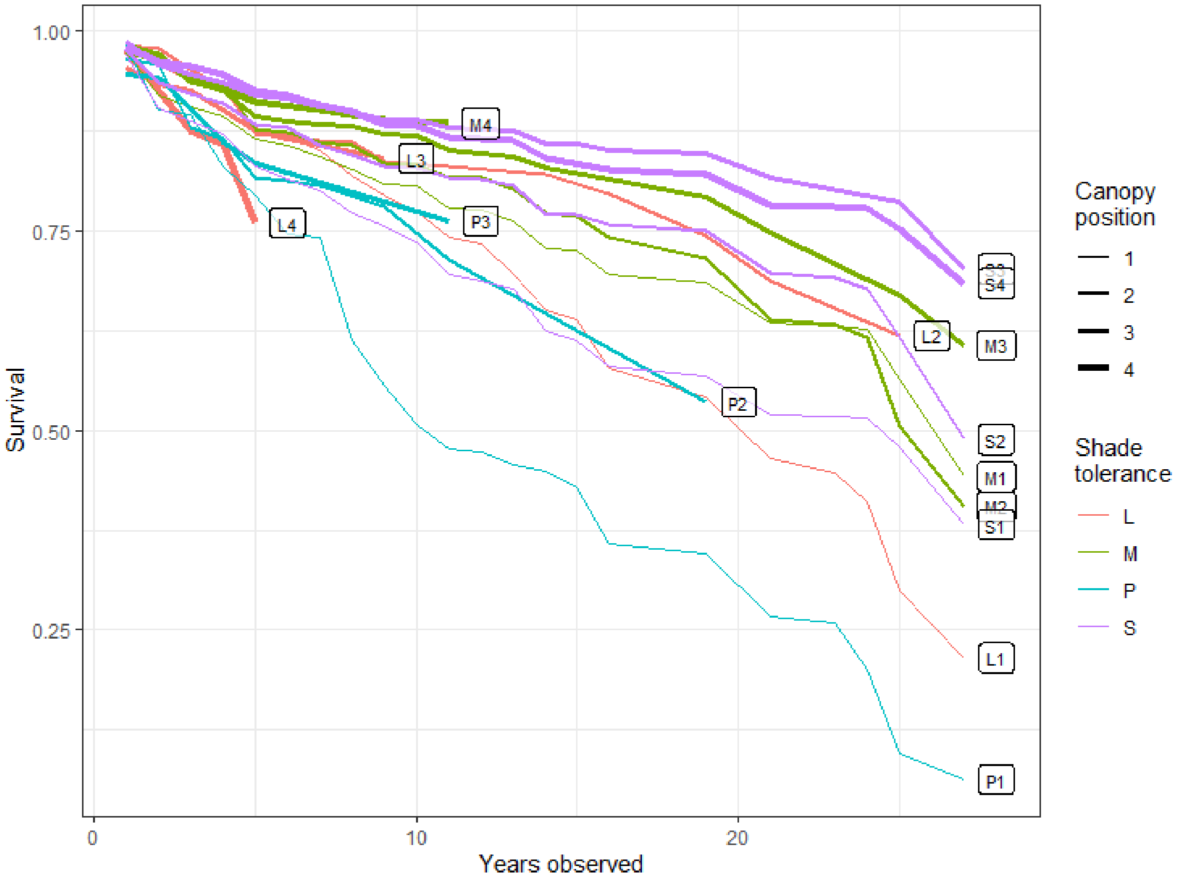
Script 4 : Workflow for survival curves and Annual Mortality Rates (AMR) by model groups

```r
# workflow for mortality functions.  This uses R package 'survival'
# wortkflow updated 25-Feb-2020

options(warn=-1)
library(tidyverse)
library(survival)
library(magrittr)
library(ggfortify)
library(Hmisc)

# add model ID and gccode
# extract required fields from treeinc and augment with model number
mort.data <- treeinc %>% inner_join(gcstats, by='gcno') %>%
  transmute(tid=row.names(.), cc=cc.x, gcno, gm, yint=as.numeric(date2-date1)/365,mort,dmg) %>% na.omit()
# tidy up yint: round down to year below except if zero, change to 1 (1st year mortality an be observed)
mort.data$yint <- with(mort.data, ifelse(yint>1, floor(yint), 1))
# mort.data has following columns:
# tid - original row name in treeinc dataframe
# cc - two letter country code
# gno - genus number in gen.nt, gen.inc etc
# mn - model ID, per inc.models$Model, GMG$Model (name changed to avoid conflict over model/Model)
# yint - interval in years over which tree observed
# mort - whether alive (0) or dead (1) at end of interval
# dmg - 0 for sound tree, 1 if any prior comment re damage or disease
# dataset is cleaned to remove any NAs in any column
# -----------------------------------

# Kaplan-Meier survival curve, all data
kmf <- survfit(Surv(yint, mort) ~ gm, data=mort.data)

# extract survfit result into a simple data frame.  For strata, strip out initial 'mn=' text with str_sub
km.data <- data.frame(mn=str_sub(summary(kmf)$strata,4), yr=summary(kmf)$time, sv=summary(kmf)$surv,
                      nt=summary(kmf)$n.risk,
                      nd=summary(kmf)$n.event)
# calculate average annual mortality up to each year
km.data %<>% mutate(amr= 1-sv^(1/yr))

# survival curves.  This is saved as Fig 3
# labels for end of lines with model ID (mn) and coordinates for label
mlabels <- km.data %>% group_by(mn) %>% summarise(yr=last(yr), sv=last(sv), amr=last(amr)) %>%
   mutate(mn_let=str_sub(mn,1,1), mn_num=str_sub(mn,2,2))
km.data$mn_let <- str_sub(km.data$mn,1,1)
km.data$mn_num <- str_sub(km.data$mn,2,2)

```

```
45
46  fig3 <- ggplot(km.data, aes(x=yr, y=sv, group=mn, color=mn_let, size=mn_num)) + geom_line() +
47          scale_size_manual(values=c(0.5, 1, 1.3, 1.6 )) + theme_bw() +
48          labs(x="Years observed", y="Survival", color="Shade\ntolerance", size="Canopy\nposition") +
49          geom_label(data=mlabels, mapping= aes(x=yr, y=sv, label=mn), color='black', nudge_x=1, alpha=0.7,
50                    size=3)
51  fig3
52
53  # the same style of chart for AMR curves
54  fig4 <- ggplot(km.data, aes(x=yr, y=amr, group=mn, color=mn_let, size=mn_num)) + geom_line() +
55     scale_size_manual(values=c(0.5, 1, 1.3, 1.6 )) + theme_bw() +
56     labs(x="Years observed", y="Annual Mortality Rate", color="Shade\ntolerance", size="Canopy\nposition") +
57     geom_label(data=mlabels, mapping= aes(x=yr, y=amr, label=mn), color='black', nudge_x=1, alpha=0.7, size=3)
58  fig4
59
60  # mean and median AMRs for each model
61  amrs <- km.data %>% group_by(mn) %>% summarise(avg=mean(amr), med=median(amr))
62
63  # add mean growth rates to inc.models
64  inc.models %<>% left_join(amrs, by=c("gm"="mn")) %>% transmute(gm, cid, ng, nt, d95, dinc, amr=avg)
```

Lines 1-10 assign the required libraries and turn off warning messages. Lines 11-26 create a working dataframe called *mort.data* which has the fields required for the Kaplan-Meier survival analysis, as documented in the comments on lines 18-24. At line 29 the function *survfit*, from library *survival*, fits the K-M curves. The result object (*kmf*) is transformed into a dataframe *km.data* with consistent column names: *mn* for model ID, *yr* as elapsed time in years, *sv* as estimated fraction of population surviving to that time, *nt* as the estimated number of trees remaining alive at that time, and *nd* as the estimated number that have died since the preceding year interval. A further column *amr* is added at line 36 with the mean AMR up to the specified year, from the relation:

Figure 3 : Kaplan-Meier survival curves fitted to data aggregated by growth model

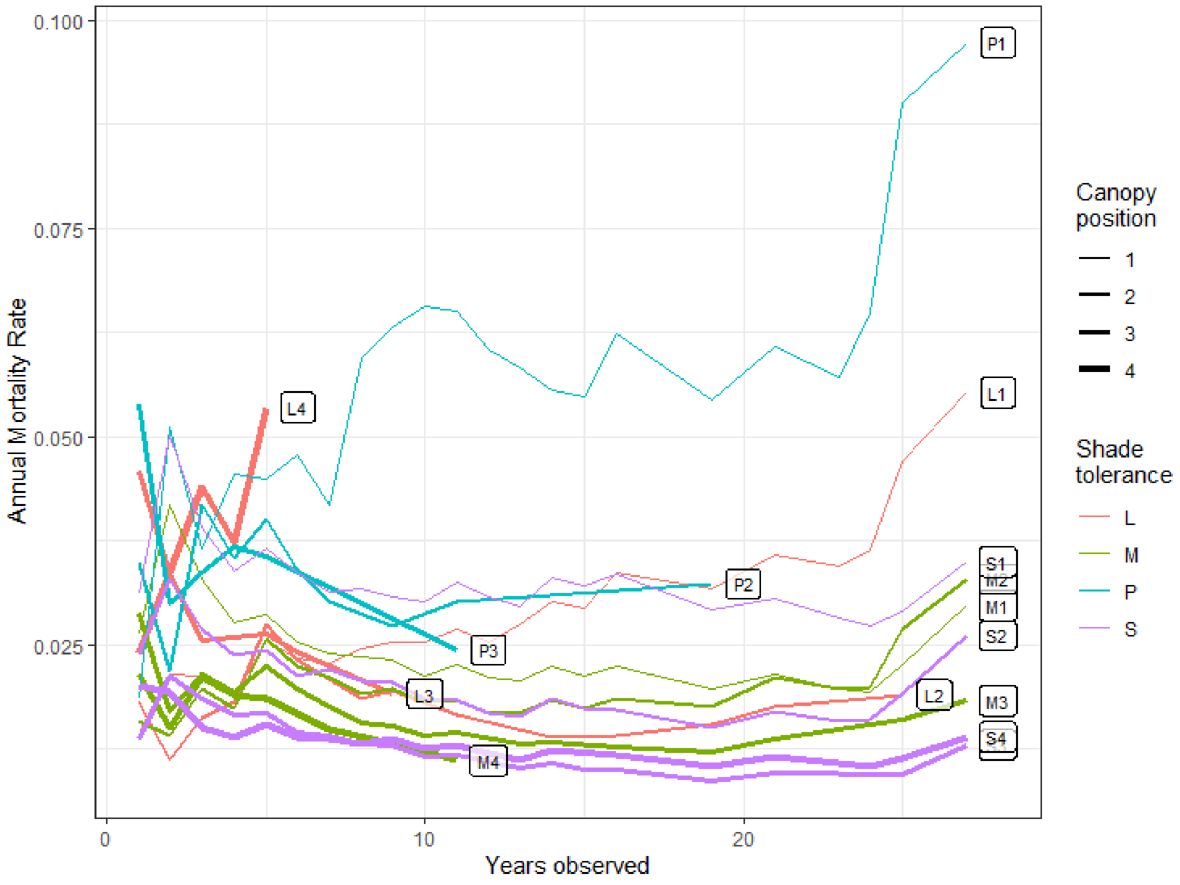


$$\text{AMR} = 1 - (\text{Survival fraction})^{1/(\text{years elapsed})}$$

At lines 40-41, a table of labels are constructed, used in Figures 3 and 4 to annotate the end of each line. This has the model ID, *mn*, and the line end coordinates for survival and mortality. Lines 42-43 derive labels for the letter and number part of each model ID for legends on the figures.

Figure 3 is created at lines 46-50, and Figure 4 at lines 54-57, exactly as reproduced here. At line 61 a table of mean and median AMRs for each growth model are constructed. These are added to the table *inc.models* for later reference at line 64.

Figure 4 : Mean mortality (AMR) rates by growth model groups



4.2. Effect of ancillary factors on mortality: Damage and dominance

In principal, and as the author has noted in other studies (eg Alder & Silva, 2000; Alder *et al.*, 2012, Alder, 1995), there is ample evidence that mortality correlates with crown position (canopy status) and with evidence of damage or disease. In this dataset, the *cpos* and *dmg* flags in table treeinc and derived tables can be used to examine this. However, because many different plots have been combined, with varying measurement protocols, some cautionary notes may be sounded:

Crown position and damage codes were not always recorded for all trees, particularly small ones or with sub-sample plot designs. Therefore, whereas a code of 1 for damage definitely indicates some indication of damage or disease, a code of zero is not necessarily a sound or healthy tree. The same applies to crown position. A code of 1 is definitely a dominant, codominant or emergent tree, but 0 does not necessarily mean that the tree is in the understorey or intermediate position.

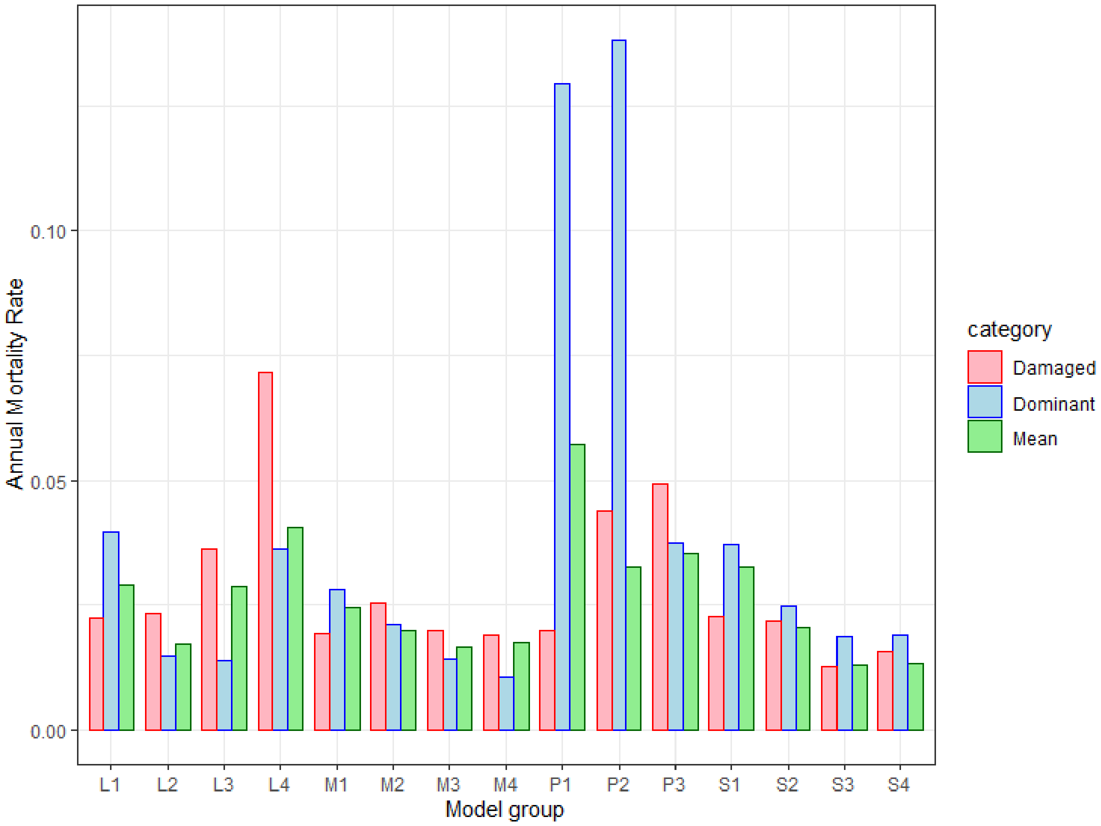
Table 3 and Figure 5 show the mean mortality rates (AMR) for the growth model groups for trees marked as dominant or emergent, and not classified as having any damage or disease (Dominant), for trees in any crown category marked or scored as having damage, decay or other signs of disease (Damaged), and the mean for all trees (Mean), including intermediate and lower canopy undamaged trees. In R script 5, Table 3 corresponds to data frame *amrs*.

Table 3 : Mortality rates for model groups by tree condition category

Model group	Damaged/ Diseased	Dominant trees	Mean, all trees
L1	2.3%	4.0%	2.9%
L2	2.3%	1.5%	1.7%
L3	3.6%	1.4%	2.9%
L4	7.2%	3.6%	4.1%
M1	1.9%	2.8%	2.5%
M2	2.5%	2.1%	2.0%
M3	2.0%	1.4%	1.7%
M4	1.9%	1.1%	1.7%
P1	2.0%	12.9%	5.7%
P2	4.4%	13.8%	3.3%
P3	4.9%	3.8%	3.5%
S1	2.3%	3.7%	3.3%
S2	2.2%	2.5%	2.0%
S3	1.3%	1.9%	1.3%
S4	1.6%	1.9%	1.3%

Figure 5 shows the same data as a bar chart, with the red bars for damaged trees, blue for dominants, and green as the overall mean, including other, lower canopy health and undamaged trees.

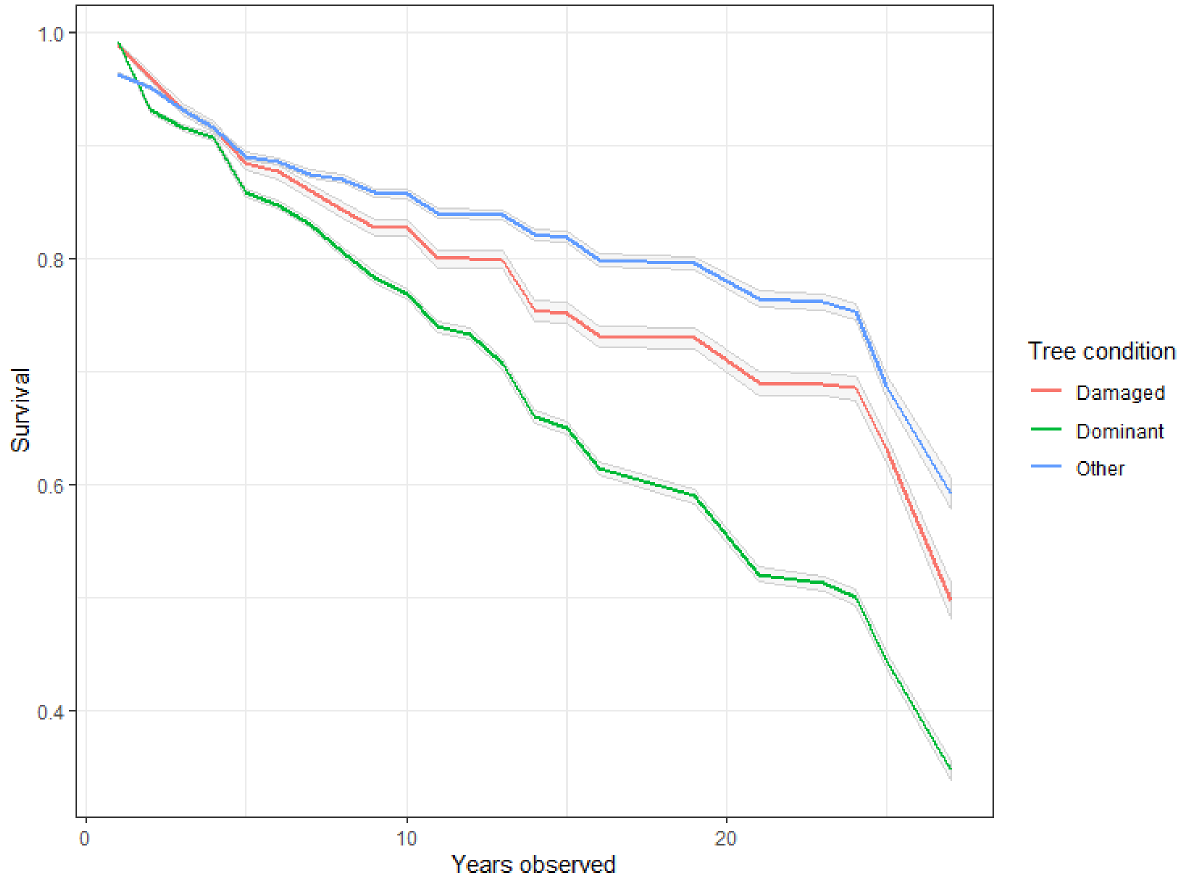
Figure 5 : Comparative mortality rates for model groups by tree condition



It was expected before doing this analysis that damaged and understorey trees would show higher mortality rates than healthy or undamaged dominants, but this is not the case. For the small, fast growing trees in the P1 and P2 groups, typically short-lived pioneer species such as Cecropia, the dominant trees have substantially higher mortality than the mean. Only for 6 of the 15 groups show the expected higher mortality for damaged trees, and for 8 groups, healthy dominants have the highest mortality.

To explore this further, and try to establish confidence intervals for the figures, Kaplan-Meier survival curves were fitted to the overall data classified by the three tree conditions: Dominant, damaged, and Other (sound lower canopy/understorey trees). The results, with 95% confidence intervals for the survival estimates as grey shading, are shown in Figure 6. For this analysis, groups P1 and P2 were excluded as they are very numerous and also anomalous, as small, short-lived pioneer-type .

Figure 6: Survival curves for pooled data, excluding groups P1 and P2, for dominant, damaged, and other trees. *95% confidence intervals as grey shading*



Here it is very clear that survival is lower for dominant and emergent crown classes than for the healthy understorey trees (<u>Other</u> category), with the damaged trees being intermediate.

The R script to produce these analyses is shown in Script 5 below. As it follows generally the techniques used in script 4 and is we3ll-commented, the details will not be described further.

Script 4 : R workflow for ancillary factors in mortality (Figures 5 and 6, Table 3)

```r
# workflow for mortality for damaged and for undamaged, dominant trees. This produces tables and a histogram,
                      with
# error bars, for average mortality rates and rates for dominant, sounds trees, and for damaged trees
# workflow updated 26-Feb-2020

# required libararies
library(tidyverse)  # dplyr, ggplot etc.
library(magrittr)   # reverse pipe operator %<>%
library(survival)   # kaplan-meier survival function (survfit)
library(ggfortify)  # extras for ggplot
library(Hmisc)      # binomial confidence intervals, binconf

# turn off warning messages
options(warn=-1)

# extract required fields from treeinc and augment with model number
mort.data <- treeinc %>% inner_join(gcstats, by='gcno') %>%
  transmute(tid=row.names(.), cc=cc.x, gcno, gm, yint=as.numeric(date2-date1)/365, mort, dmg, cpos) %>% na.omit()
# tidy up yint: round down to year below except if zero, change to 1 (1st year mortality an be observed)
mort.data$yint <- with(mort.data, ifelse(yint>1, floor(yint), 1))

# mort.data has following columns:
# tid - original row name in treeinc dataframe
# cc - two letter country code
# gno - genus number in gen.nt, gen.inc etc
# gm - model ID, per inc.models$Model, GMG$Model (name changed to avoid conflict over model/Model)
# yint - interval in years over which tree observed
# mort - whether alive (0) or dead (1) at end of interval
# dmg - 1 if any prior comment or scoring re damage or disease, 0 if no comment or not assessed (at 1st
                      measurement)
# cpos - 1 for dominant trees, 0 for intermediate/understorey or not assessed (at 1st measurement)


# ---- average mortality, all trees ----
# K-M survival curves
kmf.all <- survfit(Surv(yint, mort) ~ gm, data=mort.data)
# extract result into a simple data frame.
km.data.all <- data.frame(gm=str_sub(summary(kmf.all)$strata,4), yr=summary(kmf.all)$time,
                          sv=summary(kmf.all)$surv, nt=summary(kmf.all)$n.risk, nd=summary(kmf.all)$n.event)
# calculate average annual mortality up to each year
km.data.all %<>% mutate(amr= 1-sv^(1/yr))
# for each model, get number of trees, mean mortality rate, and equivalent number dead
amrs <- km.data.all %>% group_by(gm) %>% summarise(amr=mean(amr))

# ----- mortality rate for dominant, undamaged trees
kmf.dom <- survfit(Surv(yint, mort) ~ gm, data=filter(mort.data, cpos==1 & dmg==0))
# extract result into a simple data frame.
km.data.dom <- data.frame(gm=str_sub(summary(kmf.dom)$strata,4), yr=summary(kmf.dom)$time,
                          sv=summary(kmf.dom)$surv, nt=summary(kmf.dom)$n.risk, nd=summary(kmf.dom)$n.event)
# calculate average annual mortality up to each year
km.data.dom %<>% mutate(amr= 1-sv^(1/yr))
# add dominant mortality rate to amrs table
amrs <- km.data.dom %>% group_by(gm) %>% summarise(amr.dom=mean(amr)) %>% inner_join(amrs, by='gm')

# ----- mortality rate for damaged/unhealthy trees
kmf.dmg <- survfit(Surv(yint, mort) ~ gm, data=filter(mort.data, dmg==1))
# extract result into a simple data frame.
km.data.dmg <- data.frame(gm=str_sub(summary(kmf.dmg)$strata,4), yr=summary(kmf.dmg)$time,
                          sv=summary(kmf.dmg)$surv, nt=summary(kmf.dmg)$n.risk, nd=summary(kmf.dmg)$n.event)
# calculate average annual mortality up to each year
km.data.dmg %<>% mutate(amr= 1-sv^(1/yr))
# add dominant mortality rate to amrs table
amrs <- km.data.dmg %>% group_by(gm) %>% summarise(amr.dmg=mean(amr)) %>% inner_join(amrs, by='gm')


# add mean growth rates to inc.models
inc.models %<>% left_join(amrs, by='gm') %>% mutate(amr, amr.dom, amr.dmg)

# bar chart of AMR, damaged, overall, dominant
bars <- as.tbl(melt(amrs, id.vars=1, variable.name='category', value.name='AMR'))
bars %<>% mutate(category=recode(category, 'amr.dmg'='Damaged', 'amr.dom'='Dominant', 'amr'="Mean"))
fig5 <- ggplot(data=bars, mapping=aes(x=gm, y=AMR, fill=category, colour=category))+
   theme_bw()+
   geom_bar(stat='identity', position='dodge', width=0.8) +
   scale_fill_manual(values=c("lightpink", "lightblue", "lightgreen"), aesthetics='fill') +
   scale_fill_manual(values=c("red", "blue", "darkgreen"), aesthetics='colour') +
   labs(x="Model group", y="Annual Mortality Rate")

# overall comparison of damaged and dominant trees - create a column tc (tree condition) in mort.data table with
# categories Damaged, Dominant, Other, as a factor
mort.data %<>% mutate(tc=as.factor(ifelse(dmg==1,'Damaged',ifelse(cpos==1,'Dominant','Other'))))
# filter out the P1 and P2 groups
```

```
mort.data %<>% filter(gm !='P1' & gm != 'P2')
# K-M survival curves for Damaged, Dominant, Other
kmf.tc <- survfit(Surv(yint, mort) ~ tc, data=mort.data)
# extract result into a simple data frame.
km.data.tc <- data.frame(tc=str_sub(summary(kmf.tc)$strata,4), yr=summary(kmf.tc)$time,
                          sv=summary(kmf.tc)$surv, nt=summary(kmf.tc)$n.risk, nd=summary(kmf.tc)$n.event,
                         lcl=summary(kmf.tc)$lower, ucl=summary(kmf.tc)$upper)
# plot the survival curves with confidence intervals ()
fig6 <- ggplot(km.data.tc, aes(x=yr, y=sv, group=tc, color=tc)) +
   geom_ribbon(aes(ymin=lcl, ymax=ucl), color='lightgrey', fill='lightgrey', alpha=0.2) +
   geom_line(size=1) +
   scale_size_manual(values=c(0.5, 1, 1.5 )) + theme_bw() +
   labs(x="Years observed", y="Survival", color="Tree condition")

```

5. Assigning species to growth models

5.1. Problem definition

MYRLIN is designed, as the acronym suggests (Methods of Yield Regulation with Limited Information), for use in the situation where plans for sustainable management of a natural tropical forest have to be developed with limited detailed research information on growth rates. Typically in mixed tropical forest (MTF) there will be 100-200 species well known to local forest dwellers and forestry professionals, and perhaps 300-400 others, rarer and less well identified, that are botanically identifiable. The central purpose of the work described in this report is to provide tools that allow at least the 100-200 locally well-known species to be assigned to growth models (as listed in Table 2) using simple heuristics, or approximating methods (rules of thumb). This section (5) provides these heuristics.

5.2. Wood density and increment

The relationship between wood density and diameter increment in tropical trees has long been known (Virtucio, 1976; Enquist *et al*., 1999) and there is a large body of literature on wood density and growth rate for major timber trees. King *et al.* (2006) related wood density to growth and mortality for some 9 species of Dipterocarp. For this reason, the use of wood density as a possible heuristic in estimating increment, and together with other information, assigning growth model group[s based on static data was therefore researched.

The wood density database of Zanne et al (2009), which contains 16,468 reports from 8,412 taxa (genus or genus+species) is publicly available and was downloaded. The workflow in Script 5 shows the analysis. The wood density database is read from the downloaded Excel file (line 15) and simplified to a table containing only genus name and wood density (line 15).

Script 5 : Workflow relating wood density to growth models and diameter increment

```
 1    # wood density and growth model groups
 2    # produce histogram of mean growth models with wood density means and error bars
 3    # hopefully may be useful as a decision tool
 4
 5    # required libararies
 6    library(tidyverse)  # dplyr, ggplot etc.
 7    library(magrittr)   # reverse pipe operator %<>%
 8    library(readxl)
 9
10    # read wood density database and simplify name of column 4
11    wddb  <- read_excel('T:/FAO/data/WoodDensity/GlobalWoodDensityDatabase.xls', sheet='Data', range='A1:F16469')
12    colnames(wddb)[4] <- 'WoodDensity'
13
14    # split out genus name and wood density mean and variance into dataframe genwd
15    gn <- data.frame(genus=as.vector(str_split_fixed(wddb$Binomial, '\\s+', n=2)[,1]), density=wddb$WoodDensity,
16                          stringsAsFactors = F)
17    # dataframe with genus, count of species/observations (ns), wood density (wd) and variance (vwd)
18    genwd <- gn %>% group_by(genus) %>% summarize(ns=n(), wd=mean(density), vwd=var(density))
19    # replace NA variances (1 observation) with zeroes
20    genwd[which(is.na(genwd$vwd)),'vwd'] <- 0
21
22    # build mean and confidence limits by growth models. (needs gcstats from script 3 in environment)
23    wd.gm.stats <- genwd %>% inner_join(gcstats, by='genus') %>% group_by(gm) %>%
24       summarize(m=sum(nt), totwd=sum(wd*nt), totvar=sum(vwd*nt^2)) %>% na.omit() %>%
25       transmute(gm, nwt=m, wd=totwd/m, sdwd=sqrt(totvar/m^2)) %>%
26       mutate(ucl=wd+2*sdwd/sqrt(nwt), lcl=wd-2*sdwd/sqrt(nwt)) %>%
27       inner_join(inc.models, by='gm') %>%
28       transmute(gm, ng, d95, dinc, amr, nwt, wd, sdwd, ucl, lcl)
29
30    # wood density versus increment - regression model
31    # tested linear regression :wd.inc.reg <- with( wd.gm.stats, lm(dinc ~ wd))
32    wd.inc.reg <- wd.gm.stats %>% filter(gm!='P1') %>%  lm(log(dinc) ~ log(wd), data=.)
33    # show stats from regression
34    summary(wd.inc.reg)
```

```
35      # R-squared text with label coordinates (end of regression line)
36      r2 <-  format(summary(wd.inc.reg)$r.squared, digits=3)
37      r2txt <- paste('R^2 == ',r2 )# coefficents
38      a <- coef(wd.inc.reg)[1]
39      b <- coef(wd.inc.reg)[2]
40      # predicted values from log-log regression
41      wd.gm.stats %<>% mutate(pdinc=exp(a + b*log(wd)))
42      # positioning for Rsq label at end of line
43      lx=max(wd.gm.stats$wd)
44      ly=exp(a + b* log(lx))
45
46      # plot species groups, predicted values form regress as smoothed curve
47      fig7 <- wd.gm.stats %>% ggplot(aes(x=wd, y=dinc, size=ng, color=gm), alpha=0.5) + geom_point() + theme_bw() +
48         geom_text(aes(x=wd+0.01, y=dinc+0.01, label=gm), hjust=0, size=4) +
49         scale_size(breaks=c(5, 20, 50, 100, 200), range=c(0.1,10), name="No. genera" ) +
50         geom_smooth(aes(x=wd, y=pdinc), method='loess', color='grey40', size=2, alpha=0.5) +
51         annotate('text', x=lx, y=ly-0.1, label=r2txt, parse=T) +
52         guides(color=F) + labs(x='Wood Density (g/ml)', y='Diameter increment (cm/yr)')
53      # display chart
54      fig7
55
56      # test with D95 for comparison (displayed output only, not saved)
57      summary(with( wd.gm.stats, lm(d95 ~ wd)))
```

Lines 17-29 link this table to the MYRLIN increment and growth model data, also calculating means and confidence limits of wood density by model ID. Lines 35-45 fit a logarithmic regression to the increment-density data, and also prepare text versions of the R2 for labelling the subsequent graphic. This workflow only shows the final version of this process – several regression forms including linear and exponential where tested statistically and graphically. Also to be noted ii that after visualizing the data, group P1 (small extreme pioneers) were omitted as an outlier (note Filter at line 32). Lines 47-52 produce the plot. A quick comparison of the regression of density on D95 was run (line 57) but found to be non-significant. The results of this analysis are shown in Figure 7.

The final regression model was :

$$\log_e(\mathit{increment}) = -2.7474 \ - 3.0382 \log_e(\mathit{wood\ density})$$

with an R2 of 0.669, 12 degrees of freedom, and very high significance (P = 0.00021). Taking anti-logs, this equation becomes:

$$\mathit{Increment} = 0.06409\ \mathit{Wood\ density}^{-3.0382}$$

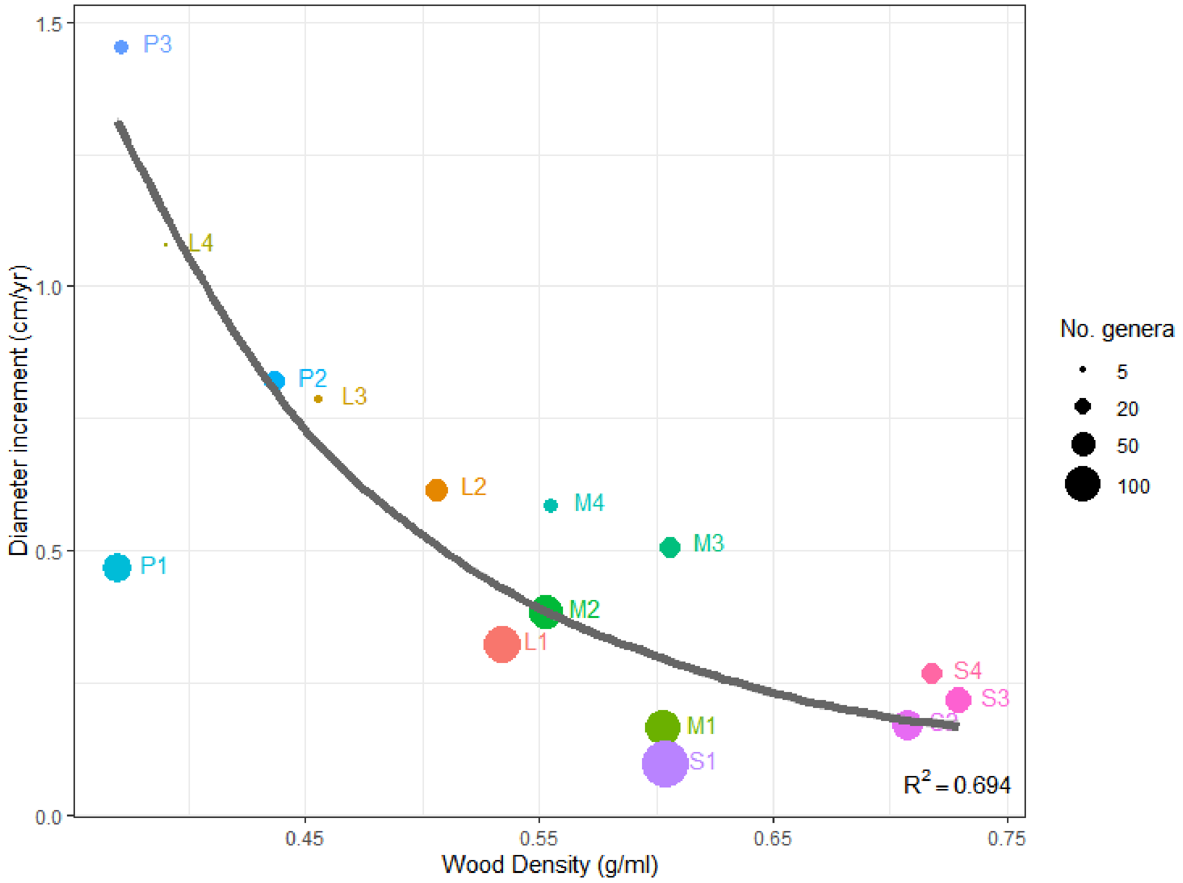
so that the cube of increment (increment^3) is inversely proportional to wood density. This is logical and as one might expect, given that one is a linear dimension and the other related to volume.

Group P1 is an outlier in terms of increment and was omitted from this regression. This is likely due to the fact that this group is dominated by data from *Cecropia* on secondary forest plots in Amazonia, observed over 27 years. After this length of time, *Cecropia* will be senescent, giving a lower mean increment for the group than might be observed over a 5-10 year time frame.

As a method, the relation between wood density and increment should not be interpreted too literally, but rather used to relate timbers known or believed to be of low, medium or high density to the corresponding low, medium and high increment groups. In actuality, density for a species can varies greatly depending on the conditions of growth of a tree, tree size and the location within the tree where density samples are taken.

Figure 7 : Relation between mean density and increment for growth model groups

(P1 is omitted from the regression as an outlier)



5.3. Direct look-up of species name

A clean table for species lookup has been compiled from the original lists in this study as saved as *splookup.rdata* (as a table) and as a text file *splookup.csv* for import into Excel or databases. The structure of this dataset is as shown in the sample listing Table 4.

Table 4 : Sample listing from *splookup* table

(Complete table has 1985 rows)

spno	cc	genus	species	nt	gm
2082	CR	*Virola*	*sebifera*	154	M2
788	BR	*Bombax*	*NA*	7	M1
678	BR	*Vochysia*	*guianensis*	12	M3
2771	CR	*Cojoba*	*valerioi*	4	M2
1912	PG	*Terminalia*	*catappa*	NA	L2
2627	CR	*Ocotea*	*atirrensis*	5	M2
174	GY	*Oxandra*	*asbeckii*	399	S1
2048	CR	*Dussia*	*macroprophyllata*	53	S2
2562	CR	*Abarema*	*macradenia*	5	M2
525	BR	*Duguetia*	*cauliflora*	3	S1

The columns are as follows:

spno .... Column of the same name in the original data sets (see Tables 2 and 3), giving species name before cleaning, and the original data for that species.

cc .... Two-letter country code.
genus .... Botanical genus, cleaned and updated name
species ..... Cleaned species name. NA implies no species given in the original data
nt .... Number of trees of this taxa in the dataset with valid increment measurements (note there may be some trees useable for mortality estimation but invalid for increment. In this case the value NA is shown).
gm........... Growth model applicable, with details that can be looked up in the *inc.models* table (Table 5).

A first step with any new species is therefore to look it up in this table. Where there are more than one entry for the same species from different geographic locales, the most similar can be used, or other information (particularly D95, estimated from inventory data), used as a basis for decision.

5.4. Decision table based on ecology and general characteristics

Table 5 provides decision rules based on mature size (D95), ecology and wood properties that allow a species to be assigned to one of the growth model groups.

Table 5 Decision table for allocating species to growth model groups

Mature diameter	**Ecology**	**Wood properties**	**Growth model**
Less than 30 cm	Pioneers, found on recently very disturbed sites, roadsides, log landings, skid trails. Not found in closed forest. Rarely live more than 20 years.	Very light, low density, white wood, decays rapidly. SG < 0.45	P2
30-40 cm			P3
Less than 30 cm.	Persistent small understorey trees and many palms.	Moderately dense, SG around 0.6	S1
30-40 cm	Persistent understorey or lower canopy trees, some larger palms.	Moderately dense, SG around 0.6.	M1
30-40 cm.	Light demanding, more persistent small pioneer trees, found in old gaps, roadsides and clearings persisting 30-40 years after disturbance. Rare in understorey of closed forest.	Less dense timber, typically light coloured, non-durable, SG around 0.5-0.55	L1
40-50 cm	Trees occurring in lower canopy and sub-canopy, typical of undisturbed forest or after long period of recovery.	Heavier, often strongly coloured timbers, SG around 0.7.	S2
40-50 cm	Lower and mid canopy trees, most common component of mature forest, suggested default group if on other indications.	Medium density wood, around SG 0.55.	M2
40-60 cm	Larger, light-demanding semi-pioneer trees common in disturbed forest after 30-40 years or in old gaps and trails, not found or rare in the understorey or lower canopy.	Lighter density and colour wood, around SG 0.5.	L2
55-75 cm	Larger upper canopy trees, but of shade tolerant species, therefore also found in lower canopy and understorey as immature trees.	Dense, often dark or coloured wood, SG >0.7.	S3
55-75 cm	Larger upper canopy trees, not common in understorey or lower canopy, but typical of recovered (40 years + post disturbance) or undisturbed forest.	Denser timber, often medium coloured, durable, SG around 0.6.	M3
60-80 cm	Large light demanding trees typically regenerating in tree-fall gaps, occurring as upper canopy trees, not found in understory or under closed canopy.	Light coloured, low density timber, SG around 0.45.	L3
80 cm +	Large emergent and upper canopy trees, but also shade tolerant and found in the understorey or sub-canopy as immature individuals, not typical of disturbed forest unless specially protected. Very long lived, often valuable timbers.	Heavy darker timber, SG > 0.7.	S4

Mature diameter	**Ecology**	**Wood properties**	**Growth model**
80 cm +	Larger emergent and upper canopy, some shade tolerance, may occur in lower canopy as regenerating individuals, default group for larger trees.	Moderate density, SG around 0.55	M4
80 cm+	Larger emergent and upper canopy trees, light demanding, gap opportunist, usually with wide spreading crown.	Light colour, low density wood, non-durable as timber, SG around 0.4.	L4

6. Conclusions

6.1. Processes and analyses completed

The analyses described in the data sets from Brazil, Guyana and Papua New Guinea, and Cost Rica, which contributed to the original MYRLIN model (Wright & Alder, 2000; Alder et al, 2002), were based on data from PSP re-measurements available until around 1998. This study has updated the system with new measurements available up to 2010-13, based on data sets used by Alder at al (2012) and Fox et al (2011).

The principal objective of the MYRLIN study was to find a robust method for projection stand growth in mixed tropical forest without having detailed PSP measurements in a given location. In this new work, the same general principal has been applied to datasets extending over a range from 1981 to 2013, as detailed in Table1 (page 5). The species lists from the various data sources have been cleaned to provide updated and corrected botanical names, using FAO tools in development (Vesa, 2020). The group centroids have been re-calculated, using weighted k-means analysis (as in the original study) to form 15 groups, represented by the letters P (Pioneer, typically fast growing species), L (non-pioneer, light demanding species), M (main series or species of median characteristic) and S (shade tolerant, slow growing species), and the numeric suffixes 1-4 according to their typical mature size, as identified by the 95% quantile of the diameter distribution (D95).

These groups are related to global wood density information from the Wood Density database of Zanne et al (2009), and provide a good and highly significant correlation (R^2 69%, n=14) with the P1 group omitted as an outlier (Figure 7, page 24). From this a table (Table 5, page 25) is given to assist in assigning species to groups when there is no information available on their growth rates. This uses information on their typical maximum size (approximate D95) from inventory data or anecdotal knowledge, their ecology, again from local knowledge of occurrence on disturbed sites, gaps, in the understorey, mid-canopy, or as emergents or upper canopy trees, and wood density, durability and colour, from local knowledge and usage or wood density databases.

6.2. Output datasets

All the analysis described have been carried out using R, apart from preliminaries with respect to the original data files that were done in SQL. All the R scripts and tables (literally, dataframes or tibbles) are listed below in Table 6, with the respective columns., and can be downloaded in the original R format as an .rdata file, or in Excel compatible tabular formats as .csv files, as listed in Table 6.

Table 6 : Output tables and related R scripts

Table name	Records	Description and column details
treeinc	128029	Original tree measurement data, with columns described in Table 2 and section 2.2 (page 6).
splist	2129	Original consolidated species lists from Brazil, Guyana, Costa Rica and Papua New Guinea, in many cases with local names, with columns shown in Table 3, as described in section 2.3.
genspp	2129	Species list with corrected/updated names and links to *splist* and *gcstats* tables, columns as follows: spno *species number in splist table* gcno *species number in gcstats table* cc *country code (BR, GY, CR or PG)* oldName *botanical name in splist (original species list)* newName *botanical name after cleaning/updating (Script 1)* genus *genus name, split from newName* species *species name, split from newname*

Table name	Records	Description and column details
		nt *corresponding number of trees in treeinc with this ID*
gcstats	1008	Species grouped into genera (as many taxa are only identified to genera) within each country, assigned a unique key number, with counts of taxa and trees, and the mean increment and D95 statistics for the genera. The table is created, and statistics calculated in R Script 2. gcno *Key : Unique number for each genus/country entry in table* gm *Growth model ID from inc.models* cid *Growth model unique number (numeric ID)* cc *Country code* genus *Genus name* ns *Number of entries for that Genus and country in genspp table* nt *Number of trees in treeinc for this Genus* dinc *Mean diameter increment for trees of this genus, from treeinc* d95 *95% diameter quantile for trees of this genus, from treeinc*
inc.models	15	Table of increment models fitted by k-means ordination of the *gcstats* data in Script 3, with annual mortality rate data added in Script 4. Columns are as follows: gm *Growth model ID* cid *Growth model serial number* ng *Number of genera (from gcstats) for this growth model* nt *Number of trees in treeinc table for this growth model* dinc *Mean diameter increment, cm/yr* d95 *Mean 95% quantile of all genera in this group* amr.dmg *Mortality rate for trees flagged as damaged in treeinc table* amr.dom *Mortality rate for trees flagged as dominant and also not damaged, in treeinc table* amr *Mortality rate for all trees in group, irrespective of damage or crown status* dbrk *Diameter at which biomass losses from mortality exceed gain from growth.*
genwd	1683	Wood density database summarised by the genera found in the *gcstats* table (see Script 5) to give the following columns: genus *Genus name. Only those found in gcstats are included.* ns *Number of entries in wood density database for the species found for the genus (some species have multiple entries)* wd *Mean wood density for species found in genus.* vwd *Variance of wood density values found*
wd.gm.stats	15	A table created in Script 5 that gives the growth model group ID, mean diameter increment and D95, together with observed wood density, upper and lower confidence limits, and predicted wood density from an allometric regression. The table gives the data underlying Figure 7. The columns in the table are: gm *Growth model ID* ng *Number of genera (from genwd) for this growth model* d95 *Mean 95% quantile of all genera in this group (from inc.models)* dinc *Mean diameter increment, cm/yr (from inc.models)* amr *Annual mortality rate (from inc.models)* nwt *Number of trees for this growth model, from treeinc* wd *Mean wood density for the growth model, from genwd* sdwd *Standard deviation of wood density, from genwd* lcl *Lower 95% confidence limit of mean wood density* ucl *Upper 95% confidence limit of mean wood density* pdinc *Predicted diameter increment from allometric regression of increment on density*
splookup	1985	A clean and minimalist species list, derived from genspp, with the following columns: spno *Species number in original list splist* genus *Genus name, cleaned and updated* species *Species name, cleaned and updated, NA if a genus-only entry.* nt *Number of trees of this species in treeinc.* gm *Growth model, for lookup in inc.models, wd.gm.stats*

These tables can all be downloaded from the link given in section 2.3 (page 6), either individually as CSV files with the extension .csv, or collectively as an R workspace, as *tables.rdata*. A complete mirror of the final R workspace, after completing the various scripts is downloadable as *workspace.rdata*. This includes the above tables, but also many other intermediate objects which are not documented here but are created in the scripts.

6.3. Next steps

FAO is currently revising and extending the original MYRLIN model, having already tested an R version, to be integrated with the Arena framework.   This should allow online, interactive use to support the development of sustainable management plans and synoptic projections. The tables and revised models provided here are a key input to this. It is hoped that the decision rules in Table 5 and the lookup table of species models can become a semi-automated heuristic to facilitate rapid calibration of a data set.

The consultant is currently undertaking further work to develop MYRLIN from the modelling perspective, and will contribute the R code and concepts from this to FAO for integration within Arena as it is developed. This will focus particularly on regeneration, recovery and successional dynamics following disturbance. The original MYRLIN model, using classical stand projection methods, is adequate in situations of careful and moderate management of tropical forest, but does not reflect well the changes that occur in situations with a high degree of disturbance, as with recovering secondary forest.

6.4. Conclusion

This report has updated the original growth data and species groups for MYRLIN, using the same datasets with remeasurements for another 10-15 years work, up to around 2010-2013, based on published sources. The datasets and models are contributing to the further development of these tools within the context of FAO's online forest planning toolkits, intended in the near future to be surfaced on the Arena framework.

References

Alder, D (1995) Growth modelling for mixed tropical forests. Department of Plant Sciences, University of Oxford. Tropical Forestry Paper 30, 231 pp.

Alder, D (2017) GEMFORM 2017 – Users Guide and Technical Notes. Technical Report, prepared under contract to van der Hout Consulting and Guyana Forestry Commission. 23 pp.

Alder, D; Oavika, F; Sanchez, M; Silva, , JNM; Van der Hout, P; Wright, HL. (2002) A comparison of species growth rates from four moist tropical forest regions using increment-size ordination. *International Forestry Review* 4(3)196-205.

Alder, D; Silva, JNM (2000) An empirical cohort model for the management of Terra Firme forests in the Brazilian Amazon. *Forest Ecology & Management*, 130:141-157.

Alder, D; Silva, JNM; Carvalho, JOP de, Lopes, J do C; Ruschel, AR (2012) The cohort-empirical modelling strategy and its application to forest management for Tapajos Forest, Para, Brazilian Amazon. *Bois et Forets des Tropiques* 314(4)17-23.

Enquist, BJ; West, GB; Charnov, EL; Brown, JH (1999) Allometric scaling of production and life-history variation in vascular plants. *Nature* 401: 6756, 907-911.

Fox J.C., Vieilledent G. and Keenan R.J. 2011. Native forest individual-tree modelling in Papua New Guinea. In '*Native forest management in Papua New Guinea: advances in assessment, modelling and decision-making*', ed. by J.C. Fox, R.J. Keenan, C.L. Brack and S. Saulei. *ACIAR Proceedings* No. 135, 106–120. Australian Centre for International Agricultural Research: Canberra.

Hawthorne, W. D. (1995) Ecological profiles of Ghanaian forest trees. *Tropical forestry papers* 29. Oxford Forestry Institute, Oxford. 345 pp.

Kaplan, EL; Meier, P (1958) Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association* **53**(282)457-481.

King, DA; Davies, SJ; Tan, S; Noor, NSMD (2006) The role of wood density and stem support costs in the growth and mortality of tropical trees. *Journal of Ecology* 94, 670–680.

McDowell, N *et al.* (2018) Drivers and mechanisms of tree mortality in moist tropical forests. Tansley Review, *New Phytologist* (2018) doi: 10.1111/nph.15027. 19 pp.

Perez, JG (2017) The use of global biodiversity databases to increase taxonomic quality in forest inventories. Presentation at UN-REDD workshop, FAO, 7 September 2017. http://tinyurl.com/r5muwz8

Silva, JNM; De Carvalho, JOP; Lopes, J. de CA; De Oliveira, RP; De Oliveira, LC (1996) Growth and yield studies in the Tapajos region, Central Brazilian Amazon. *Commonwealth Forestry Review* 75(4)325-329.

Silva, JNM; de Carvalho, JOP; Lopes, JCA; Almeida BF; Costa DHM; de Oliveira, LC; Vanclay, JK; Skovsgaard, JP (1995) Growth and yield of a tropical rainforest in the Brazilian Amazon 13 years after logging. *Forest Ecology and Management*, 71:267-274.

van der Hout, P (1999) Reduced impact logging in the tropical rainforest of Guyana. *Tropenbos-Guyana Series* 6, 335 pp.

van der Hout, P (2000a) Testing the applicability of reduced impact logging in greenheart forest in Guyana. *International Forestry Review* 2 (1) 24-32

van der Hout, P (2000b) Pibiri permanent plots: Objectives, design and database management. Tropenbos Guyana Reports 2000-2.

Vesa, L (2019) Shiny (MYRLIN) tool: draft Concept note. 29 August 2019 and following. Email communication, lauri.vesa@fao.org.

Vesa, L (2020) Species Names Validation Tool. R-script from lauri.vesa@fao.org, FAO, Rome. version 12th Feb. 2020 ver 2. Unpublished.

Virtucio, FD (1976) Tree growth as a function of specific gravity. *Philippine Forest Research Journal* 1(1)9-14.

Wikipedia (2020) Kaplan–Meier estimator. https://en.wikipedia.org/wiki/Kaplan%E2%80%93Meier_estimator

Wright, HL; Alder, D [Ed.] (2000) Humid and semi-humid tropical forest yield regulation with minimal data. Workshop proceedings, CATIE, Costa Rica, 5-9 July 1999. *Oxford Forestry Institute Occasional Papers* 52. 95 pp.

Zanne, AE; Lopez-Gonzalez, G; Coomes, DA; Ilic, J; Jansen, S; Lewis, SL; Miller, RB; Swenson, NG; Wiemann, MC; Chave, J (2009) Global wood density database. https://datadryad.org/stash/dataset/doi:10.5061/dryad.234