

Fiche TD avec le logiciel  : MASTREE

Importation des données MASTREE+

P^r Jean R. LOBRY

Ce jeu de données sous licence libre CC-BY-4.0 contient des séries temporelles sur l'intensité de l'investissement dans la reproduction pour des végétaux sis sur six continents. Les variables disponibles sont décrites en français en annexe de cette fiche.

Contents

1	Licence et accès	2
2	Format des données	3
3	Annexe : liste des variables	4
	Références	35

1 Licence et accès

1.1 Licence

LES DONNÉES de MASTREE+ sont publiées sous la licence libre CC-BY-4.0 qui permet aux utilisateurs de les copier et redistribuer, de les adapter et de les modifier, sur tout support et tout format, et ce dans n'importe quel but, y compris lucratif. Vous devez citer la publication associée [8] et fournir un lien vers la licence et indiquer si elle a été modifiée¹.

LES auteurs utilisant les données RENECOFOR² doivent remercier le réseau RENECOFOR et envoyer une copie de la publication³. Les auteurs utilisant les données du parc national de Lopé au Gabon⁴ doivent citer le jeu de données originel⁵, remercier l'Agence Nationale des Parcs Nationaux (ANPN) et l'université de Stirling et envoyer une copie de la publication⁶.

B IEN QUE cela ne soit pas explicitement requis, il est évident que toute publication exploitant une source de données particulière doit citer la référence originelle. Elles sont disponibles dans l'annexe S4 de [8]. La colonne `Study_ID` permet de tracer la source originelle des données :

- 1° M* données extraites de la littérature, voir la référence correspondante dans l'annexe S4.
- 2° A* données issues des compilations d'ASCOLI et collaborateurs [2, 3].
- 3° P* données issues de la compilation de PEARSE et collaborateurs [14].
- 4° D* données non publiées avant [8].

1.2 Accès

UNE première possibilité consiste à récupérer le fichier⁷ sur le site de l'éditeur⁸ de l'article [8]. C'est un fichier au format CSV, les colonnes sont séparées par des virgules, de 18 254 364 octets. C'est la version originelle des données déposée avec la publication, donc pas forcément à jour.

UNE seconde possibilité est de passer par Dryad⁹ mais il semble que le DOI n'ait pas encore été activé au moment de l'écriture de ce document (2022-03-18).

LES auteurs [8] recommandent d'utiliser le dépôt GitHub¹⁰ pour bénéficier de la dernière version du jeu de données. Un autre intérêt est que les données sont disponibles au format XDR [15]. C'est un format binaire, beaucoup plus compact (491 850 octets sur disque dans la version utilisée ci-après) donc très rapide à lire, compatible multi-plateformes, directement importable sous . Dans le code ci-après l'instruction pour télécharger le fichier (`download.file()`)

¹Voir <https://creativecommons.org/licenses/by/4.0/>

²Reference = RENECOFOR_2020

³manuel.nicolas@onf.fr

⁴Reference = Bush_2021

⁵<http://hdl.handle.net/11667/152>

⁶science@parcsgabon.ga et k.a.abernethy@stir.ac.uk

⁷gcb16130-sup-0001-appendixs1.csv

⁸<https://onlinelibrary.wiley.com/doi/full/10.1111/gcb.16130>

⁹<https://doi.org/10.5061/dryad.18931zd02>

¹⁰<https://github.com/JJFoest/MASTREEplus>

est mise en commentaire pour ne pas avoir à l'exécuter chaque fois que je compile le présent document.

```
path <- "https://github.com/JJFoest/MASTREEplus/raw/main/Data/"
fname <- "MASTREEplus_2022-02-03_V1.RData"
#
# download.file(paste0(path, fname), destfile = fname) # à dé-commenter
#
load(fname)
dim(mastree)
[1] 73828    32
```

NOTONS enfin qu'il existe une application Shiny¹¹ qui permet d'explorer de façon très intuitive et interactive les données disponibles.

1.3 Documents complémentaires

LES NOMS des fichiers supplémentaires de [8] sont sous une forme standard¹² où `n` peut prendre les valeurs de 1 à 6 et `ext` peut être `csv` ou `docx`.

S1 : c'est un fichier au format CSV, les colonnes sont séparées par des virgules.

S2 : description de la méthodologie pour l'orpaillage bibliographique.

S3 : description de la méthodologie pour le contrôle de qualité des données.

S4 : liste des références bibliographiques.

S5 : description des variables.

2 Format des données

2.1 La classe de `mastree`

L'OBJET `mastree` est de la classe `tibble`. Si l'on préfère travailler avec des objets de la classe `data.frame` plus classiques, il suffit de faire :

```
mastree <- as.data.frame(mastree)
```

2.2 Clef d'identification

D'APRÈS les auteurs [8] chaque série temporelle est univoquement identifiée en combinant `Alpha_Number`, `Site_number`, `Variable_number` et `Species_code`. Vérifions qu'il n'y a pas d'années dupliquées dans les séries temporelles ainsi définies :

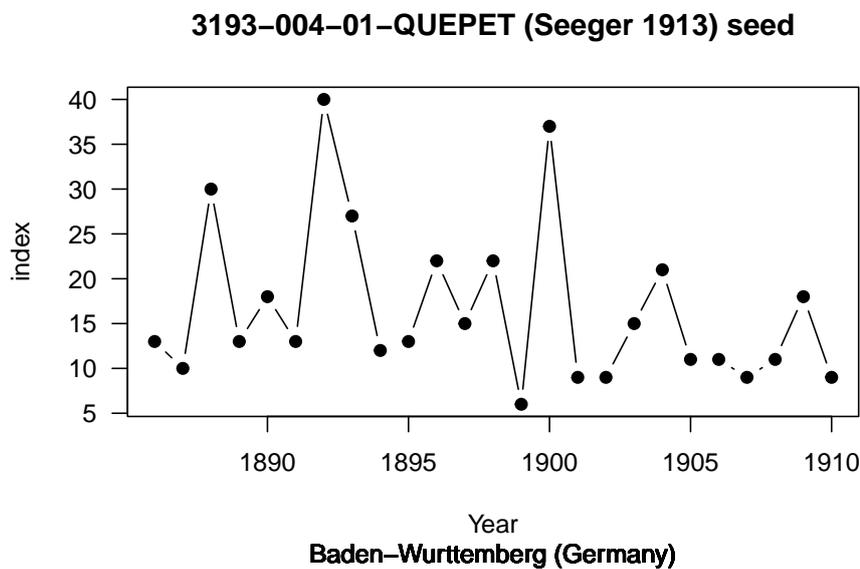
```
target <- c("Alpha_Number", "Site_number", "Variable_number", "Species_code")
mastree$ID <- apply(mastree[, target], 1, \(x) paste(x, collapse = "-"))
any(tapply(mastree$Year, mastree$ID, \(x) any(duplicated(x)))) # FALSE, OK
[1] FALSE
```

¹¹<https://mastreeplus.shinyapps.io/mastreeplus/>

¹²`gcb16130-sup-000[n]-appendix[n].[ext]`

2.3 Visualisation d'une série temporelle

```
plotID <- function(the_ID, las = 1, type = "b", pch = 19, ...){
  df <- mastree[mastree$ID == the_ID, ]
  x <- df$Year
  y <- df$Value
  main <- paste(the_ID, " (", unique(df$Reference), ") ",
               unique(df$Variable), sep = "")
  plot(x, y, main = main, xlab = "Year", ylab = unique(df$Unit),
       las = las, type = type, pch = pch, ...)
  title(sub = paste(df$Site, " (", df$Country, ") ", sep = ""))
}
plotID("3193-004-01-QUEPET")
```



3 Annexe : liste des variables

Les variables sont décrites dans l'annexe S5 de [8].

```
names(mastree)
[1] "Alpha_Number"      "Segment"          "Site_number"
[4] "Variable_number"  "Year"             "Species"
[7] "Species_code"     "Mono_Poly"       "Value"
[10] "VarType"          "Max_value"       "Unit"
[13] "Variable"         "Collection_method" "Latitude"
[16] "Longitude"       "Coordinate_flag" "Site"
[19] "Country"         "Elevation"       "Spatial_unit"
[22] "No_indivs"       "Start"           "End"
[25] "Length"          "Reference"       "Record_type"
[28] "ID_enterer"     "Date_entry"     "Note on data location"
[31] "Comments"        "Study_ID"       "ID"
```

3.1 Alpha_Number*

C'EST un code numérique de 4 caractères associé avec chaque source originelle des données et qui pointe vers la liste des références de l'annexe 4 de [8].

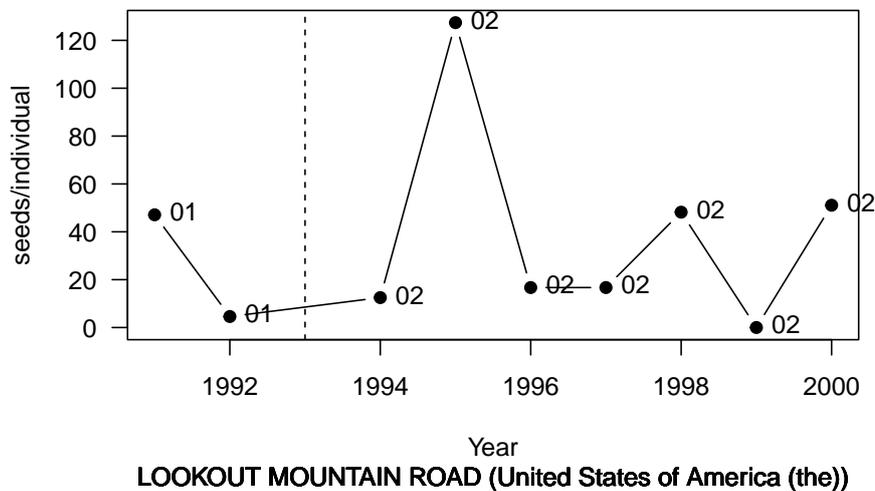
```
all(nchar(mastree$Alpha_num) == 4L) # TRUE, OK
[1] TRUE
length(unique(mastree$Alpha_Number)) # 730 in this version
[1] 730
```

3.2 Segment

C'EST un code numérique de 2 caractères utilisé pour coder les séries temporelle comportant des lacunes. Les années avec données manquantes ne sont pas explicitement codées. Par exemple, les données de Kaye *et al.*¹³ (Alpha_Number = 5087) courent de 1991 à 2000 mais avec une donnée manquante en 1993. Par conséquent les années 1991-1992 forment le segment 01 et les années 1994-2000 le segment 02 :

```
all(nchar(mastree$Segment) == 2L) # TRUE, OK
[1] TRUE
the_ID <- "5087-001-01-PYRRAD"
plotID(the_ID)
with(mastree[mastree$ID == the_ID, ], {
  text(Year, Value, Segment, pos = 4, xpd = NA)
})
abline(v = 1993, lty = 2)
```

5087-001-01-PYRRAD (KAYE 2001) seed

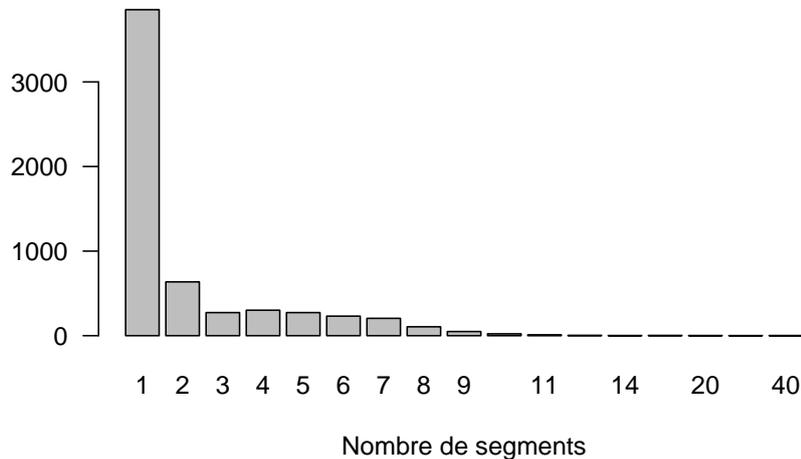


La plupart des séries sont d'un seul tenant :

```
nseg <- with(mastree, tapply(Segment, ID, \(x) length(unique(x))))
barplot(table(nseg), las = 1, main = "Fragmentation des séries",
  xlab = "Nombre de segments")
```

¹³T. N. Kaye, "Population monitoring and preliminary viability model of Snake River goldenweed, *Haplopappus radiatus*: tenth year progress report" (Bureau of Land Management Vale District and Institute for Applied Ecology, Corvallis, OR, 2001).

Fragmentation des séries



3.3 Site_number*

C'EST un code numérique de 3 caractères utilisé pour différencier les sites multiples d'une même source. C'est en fait un codage numérique de la variable Site pour une série temporelle donnée.

```
all(nchar(mastree$Site_number) == 3L) # TRUE, OK
[1] TRUE
nds <- tapply(mastree$Site_number, mastree$Alpha_Number, \(x) length(unique(x)))
summary(nds)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
1.000  1.000   1.000  4.063  2.000 428.000
head(rev(sort(nds)), n = 10)
6001 5164 3002 5181 3122 3198 3036 3048 3176 2553
428  138  129  75  75  70  64  62  54  51
```

Voici le top-10 des études ayant le plus de sites :

- 6001** Department of Forest Genetics, Federal Research and Training Centre for Forests, Natural Hazards and Landscape (BFW).
- 5164** K. Tallqvist, Results of long-time measurements of the quality of flowering and seed crop of trees. Folia Forestali 364, 1 (1978).
- 3002** Prognosen der Waldsamenernte. Allgemeine Forstzeitschrift.
- 5181** Référence manquante.
- 3122** Forschungsanstalt für Wald, Schnee und Landschaft WSL http://www.wsl.ch/mastweb/resultate/index_DE
- 3198** Slovenian Forest Service (Zavod za Gozdovje Slovenije). 2016. Poročilo zavoda za gozdove slovenije o gozdovih. Za leto 2015. Ljubljana.

- 3036** Corpo Forestale dello Stato, Ministero delle Politiche Agricole, Alimentari e Forestali, Ufficio territoriale per la Biodiversità di Verona - CNBF di Peri, Italiano
- 3048** EAN - European Aerobiological Network - Austria
- 3176** Regierungspräsidium Freiburg
- 2553** Wright, M. C., P. van Mantgem, N. L. Stephenson, A. J. Das, and J. E. Keeley. 2021. Seed production patterns of surviving Sierra Nevada conifers show minimal change following drought. *Forest Ecology and Management* 480:21.

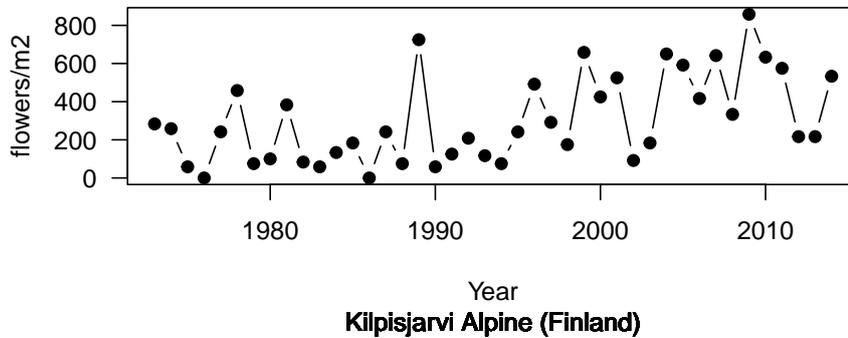
3.4 Variable_number*

C'EST un code numérique d'un caractère utilisé dans les rares cas où plusieurs mesures différentes de l'effort reproductif ont été faites. Par exemple, Boulanger-Lapointe et collaborateurs¹⁴ (`Alpha_number = 0187`) ont suivi l'effort de production de fleurs *et* de fruits pour *Vaccinium myrtillus* sur les mêmes sites.

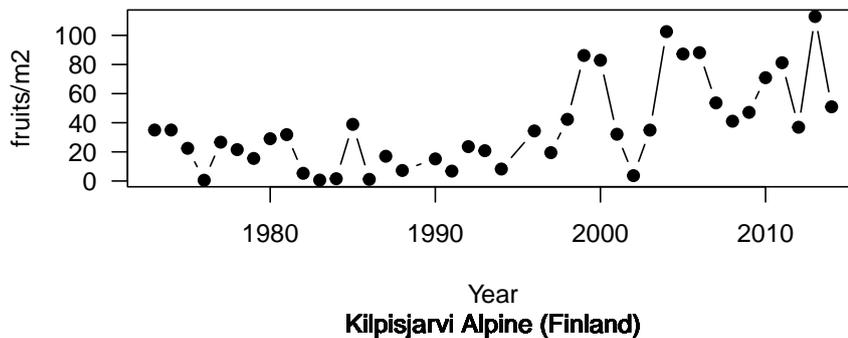
```
par(mfrow = c(2, 1))
plotID("0187-001-01-VACMYR")
plotID("0187-001-02-VACMYR")
```

¹⁴Boulanger-Lapointe, N., A. Järvinen, R. Partanen, and T. M. Herrmann. 2017. Climate and herbivore influence on *Vaccinium myrtillus* over the last 40 years in northwest Lapland, Finland. *Ecosphere* 8.

0187-001-01-VACMYR (Boulanger-Lapointe_2017) flower



0187-001-02-VACMYR (Boulanger-Lapointe_2017) fruit



B IEN QUE rares, ces études sont importantes car elles permettent d'appréhender de façon plus fine l'intensité de l'effort dans la reproduction.

```
nvar <- with(mastree, tapply(Variable_number, Alpha_Number, \(x) length(unique(x))))
(rares <- names(nvar[nvar > 1]))

[1] "0006" "0187" "0758" "0807" "0830" "0982" "1010" "1087" "1322" "1472" "1500"
[12] "1516" "1787" "1822" "1893" "2056" "2492" "2503" "2505" "3002" "3014" "3036"
[23] "3043" "3052" "3112" "3122" "3146" "3149" "3160" "3169" "3190" "3193" "3198"
[34] "3201" "3202" "4001" "6194"

cols <- c("Alpha_Number", "Species", "Country")
res <- mastree[!duplicated(mastree$Alpha_Number) & mastree$Alpha_Number %in% rares, cols]
res[order(res$Species), ]

Alpha_Number Species Country
6854 2503 Abies cephalonica Greece
4103 1500 Adansonia digitata South Africa
70756 6194 Aeschynanthus acuminatus Taiwan (Province of China)
2322 0758 Alphitonia excelsa Philippines (the)
3230 1087 Cryptomeria japonica Japan
6773 2492 Dicymbe corymbosa Guyana
5302 1893 Fagus crenata Japan
13324 3002 Fagus sylvatica Germany
15176 3014 Fagus sylvatica Germany
17029 3036 Fagus sylvatica Italy
17498 3043 Fagus sylvatica Germany
18881 3052 Fagus sylvatica Germany
21245 3112 Fagus sylvatica Austria
21392 3122 Fagus sylvatica Switzerland
22697 3149 Fagus sylvatica Belgium
23005 3160 Fagus sylvatica Netherlands (the)
25093 3190 Fagus sylvatica Poland
25361 3193 Fagus sylvatica Germany
```

25609	3198	Fagus sylvatica	Slovenia
26155	3201	Fagus sylvatica	Bosnia and Herzegovina
26185	3202	Fagus sylvatica	Germany
3101	1010	Lagerstroemia speciosa	India
22676	3146	Picea abies	Austria
23216	3169	Picea abies	Italy
27387	4001	Picea glauca	United States of America (the)
3646	1322	Pinus edulis	United States of America (the)
4126	1516	Pinus koraiensis	China
5188	1822	Pinus koraiensis	China
4018	1472	Pinus pinea	Spain
5865	2056	Pinus resinosa	Canada
2612	0830	Quercus acuta	Japan
3085	0982	Quercus alba	United States of America (the)
6897	2505	Quercus ilex	Spain
64	0006	Sasa veitchii	Japan
2419	0807	Sclerocarya birrea	South Africa
314	0187	Vaccinium myrtillus	Finland
5160	1787	Virola surinamensis	Costa Rica

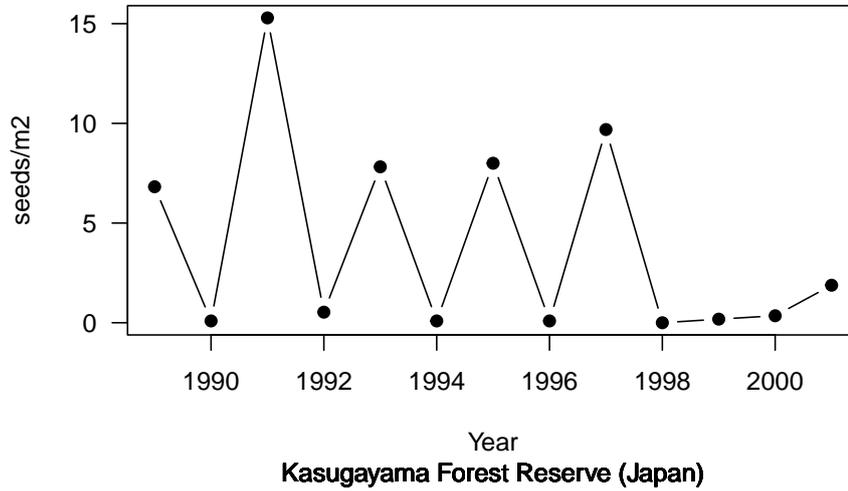
POUR ce qui est du chêne, on ne trouve aucune étude pour le chêne sessile ou pédonculé. Les 3 références pour les autres espèces de chêne sont les suivantes :

- 0830** Hirayama, D., T. Fujii, S. Nanami, A. Itoh, and T. Yamakura. 2012. Two-year cycles of synchronous acorn and leaf production in biennial-fruited evergreen oaks of subgenus Cyclobalanopsis (*Quercus*, Fagaceae). *Ecological Research* 27:1059-1068.
- 0982** Kellner, K. F., J. K. Riegel, N. I. Lichti, and R. K. Swihart. 2013a. Oak mast production and animal impacts on acorn survival in the Central Hardwoods. *The Hardwood Ecosystem Experiment: a framework for studying responses to forest management* Edited by RK Swihart, MR Saunders, RA Kalb, S. Haulton, and CH Michler. US Dep. Agric. For. Serv. Gen. Tech. Rep. No. NRSP-108:176-190.
- 2505** Carbonero, M.D., Fernández-Rebollo, P. 2014. Dehesas de encinas. Influencia de la meteorología en la producción de bellotas. *Ecosistemas*, 23, 55-63

POUR l'étude japonaise, je pense qu'il y a une typo dans MASTREE+ puisqu'il s'agit toujours de la variable `seed` mais pour trois espèces différentes (*Quercus acuta*, *Q. salicina* et *Q. sessilifolia*).

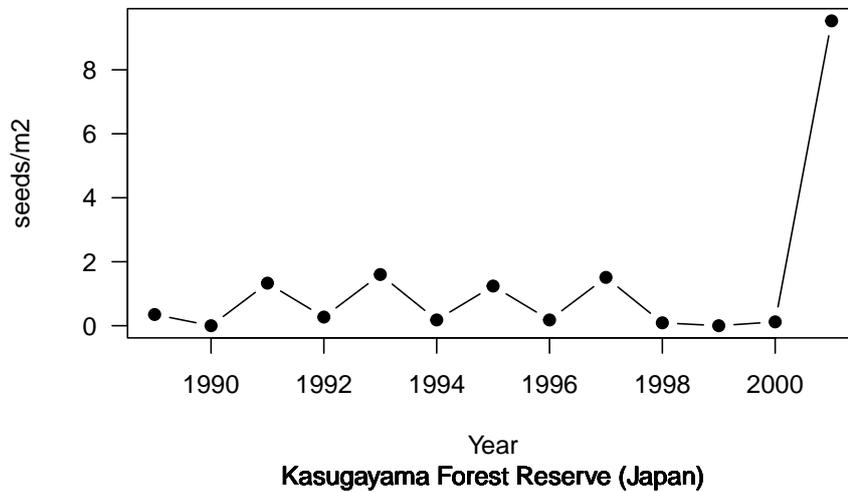
`plotID("0830-001-01-QUEACA")`

0830-001-01-QUEACA (Hirayama_2012) seed

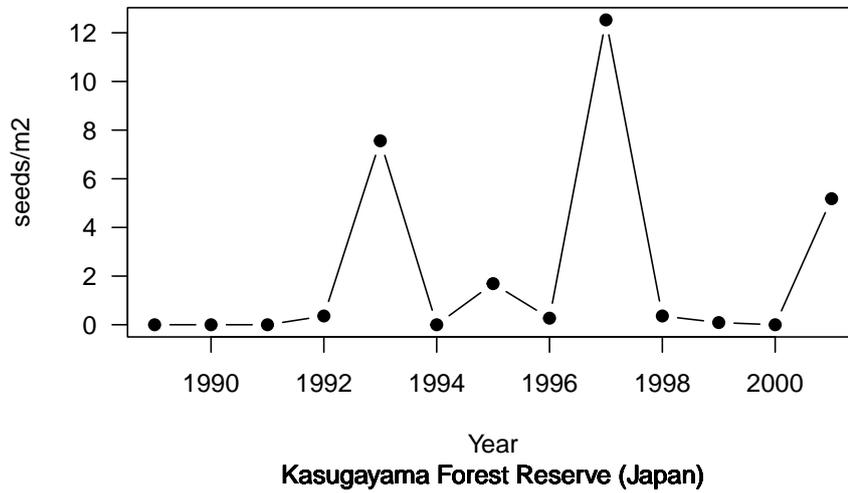


plotID("0830-001-02-QUESAL")

0830-001-02-QUESAL (Hirayama_2012) seed



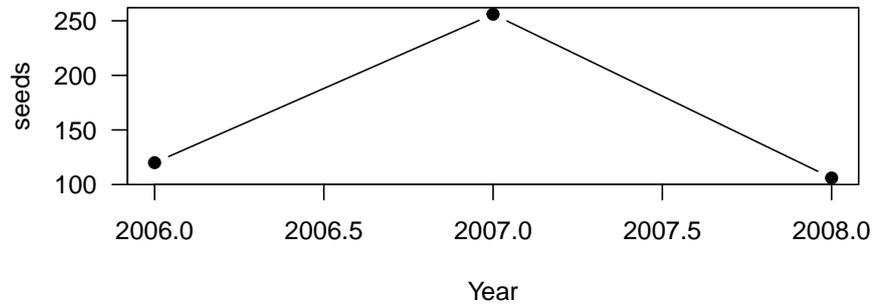
plotID("0830-001-03-QUESES")

0830-001-03-QUESES (Hirayama_2012) seed

POUR l'étude américaine, je pense qu'il y a également une typo dans MASTREE+ puisqu'il s'agit toujours de la variable `fruit` mais pour deux espèces différentes (*Q. alba* et *Q. velutina*).

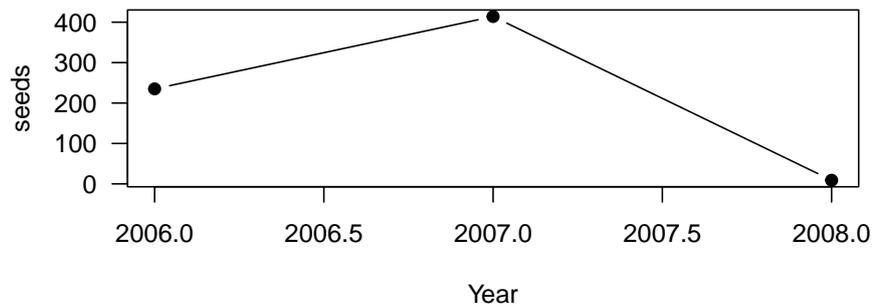
```
par(mfrow = c(2, 1))  
plotID("0982-001-01-QEALB")  
plotID("0982-001-02-QEVEL")
```

0982-001-01-QUEALB (Kellner_2013) fruit



» State Forest, Yellowwood State Forest, and Brown County State Park (United States)

0982-001-02-QUEVEL (Kellner_2013) fruit

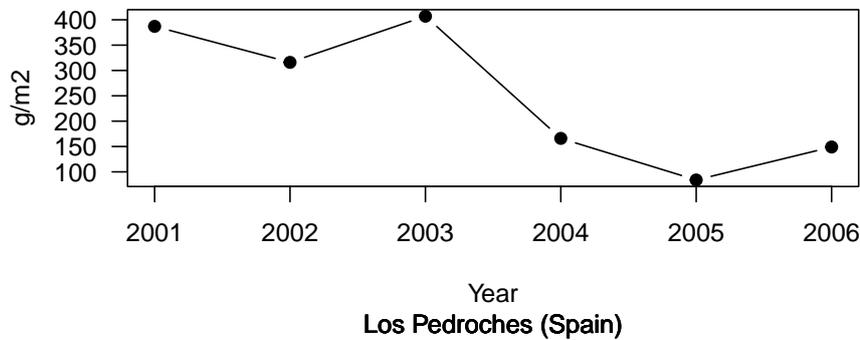


» State Forest, Yellowwood State Forest, and Brown County State Park (United States)

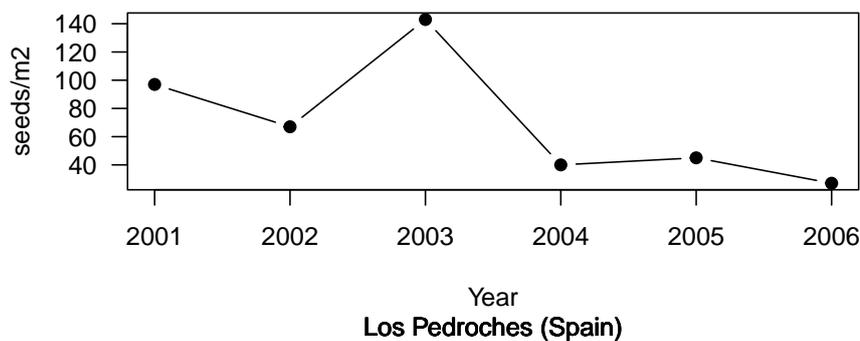
POUR l'étude espagnole il n'y a bien qu'une espèce, *Q. ilex*, mais il s'agit toujours de la variable `seed`, mais mesurée ici avec deux méthodes différentes.

```
par(mfrow = c(2, 1))
plotID("2505-001-01-QUEILE")
plotID("2505-001-02-QUEILE")
```

2505-001-01-QUEILE (Carbonero_2014) seed



2505-001-02-QUEILE (Carbonero_2014) seed



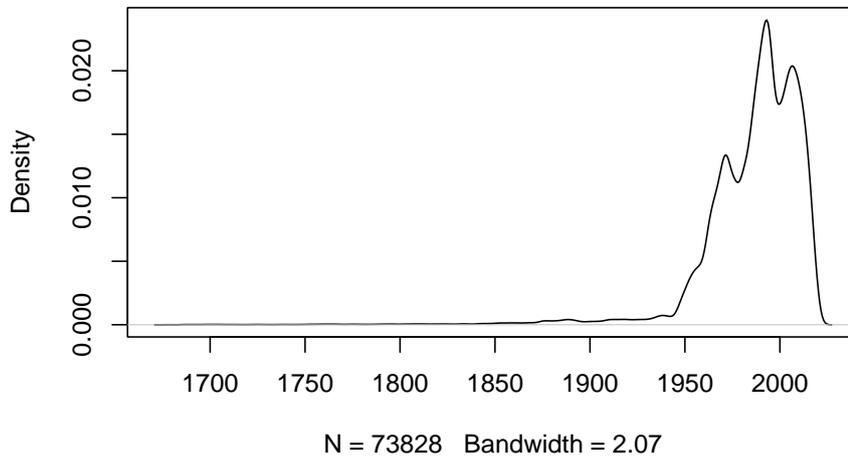
3.5 Year

COMME on pourrait s'en douter, il s'agit de l'année de l'observation codée sur 4 caractères, mais avec une subtilité pour que cela ait plus de sens biologique :

- 1° Pour les espèces qui ont une saison phénologique bien marquée, donc hors des tropiques, c'est l'année calendaire du moment de la floraison ou de la fructification.
- 2° Pour les espèces tropicales, ou sans saison phénologique particulière, la production annuelle est la somme sur une fenêtre temporelle de 12 mois choisie pour faire le plus de sens possible d'un point de vue biologique. Cette information est portée dans la variable `Comment`.

```
summary(mastree$Year)
  Min. 1st Qu. Median   Mean 3rd Qu.   Max.
  1677  1975   1991   1986  2004   2021
plot(density(mastree$Year))
```

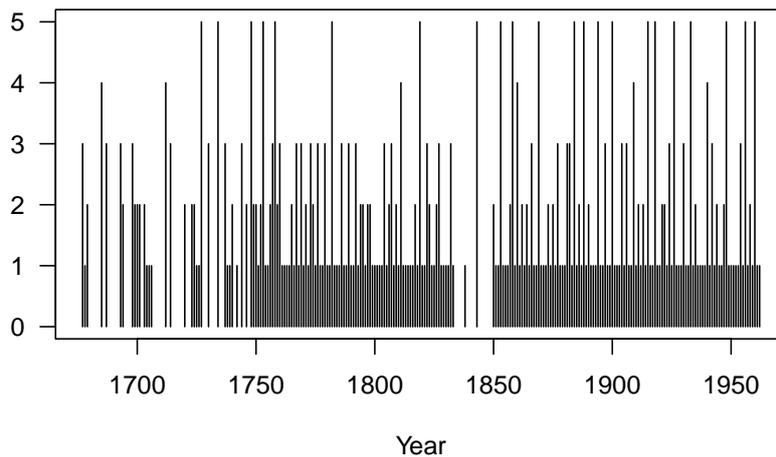
density.default(x = mastree\$Year)



LA PLUPART des données sont postérieures à 1950. La moitié des données sont des séries contemporaines entre 1975 et 2004. Par curiosité on regarde à quoi correspond la série qui commence en 1677 :

```
(vieux <- unique(mastree[mastree$Start == 1677, "ID"]))
[1] "3067-001-01-FAGSYL"
plotID(vieux, type = "h", ylim = c(0, 5))
```

3067-001-01-FAGSYL (Hase 1964) seed



Schleswig-Holstein (Germany)

C'EST une étude de W. HASE¹⁵ de suivi au long cours de la production de graines par le hêtre commun (*Fagus sylvatica*). La mesure est une variable

¹⁵Hase, W. 1964. Die Buchenmast in Schleswig-Holstein und ihre Abhängigkeit von der Witterung. Mitt. Deutsch. Wetterdienst. 31: 31/3-31/45

qualitative ordonnée à 5 modalités.

3.6 Species

Le nom des espèces suit la *Plant List nomenclature* standardisée avec le paquet  `Taxonstand` [7]. *Genus spp.* correspond à une série où la reproduction n'est identifiée qu'au niveau du genre *Genus*. « MIXED » indique une communauté de plusieurs espèces.

```
ntspersp <- with(mastree[!duplicated(mastree$ID), ], table(Species))
length(ntspersp) # 1004
[1] 1004
ptpersp <- rev(sort(table(mastree$Species)))
head(ptpersp, n = 10)
      Picea abies  Fagus sylvatica Pinus sylvestris  Larix decidua  Abies alba
      12002      11113      5288      5025      4672
Quercus robur  Quercus petraea  Picea glauca  Quercus spp.  Abies balsamea
      1605      1436      1274      1232      1097
sum(ptpersp[1:10])/sum(ptpersp)
[1] 0.6060573
```

Il y a donc de l'ordre de 10^3 espèces documentées dans cette version de MASTREE+. Pour le top-10 des espèces les mieux documentées on a entre 10^3 et 10^4 données disponibles. Le top-10 des espèces pour lesquelles on dispose du plus de données est le suivant :

- 1° *Picea abies*, l'épicéa commun, de la famille des *Pinaceae*, localisé en Europe septentrionale, le sapin de Noël par excellence.
- 2° *Fagus sylvatica*, le hêtre commun, de la famille des *Fagaceae*, localisé en Europe.
- 3° *Pinus sylvestris*, le pin sylvestre, de la famille des *Pinaceae*, localisé en Europe moyenne et septentrionale, d'Asie du Nord jusqu'en Sibérie orientale, à très grande amplitude écologique et climatique.
- 4° *Larix decidua*, le mélèze commun, de la famille des *Pinaceae*, localisé en Europe centrale et pionnière en altitude.
- 5° *Abies alba*, le sapin commun, de la famille des *Pinaceae*, localisé en Europe méridionale, occidentale et centrale.
- 6° *Quercus robur*, le chêne pédonculé, de la famille des *Fagaceae*, localisé en Europe.
- 7° *Quercus petraea*, le chêne sessile, de la famille des *Fagaceae*, localisé en Europe.
- 8° *Picea glauca*, l'épinette blanche, de la famille des *Pinaceae*, localisée en Amérique du Nord.
- 9° *Quercus spp.*, genre de la famille des *Fagaceae*.
- 10° *Abies balsamea*, le sapin baumier, de la famille des *Pinaceae*, localisé en Amérique du Nord.

IL y a un gros biais d'échantillonnage taxonomique puisque les *Pinaceae* et *Fagaceae* d'Europe et d'Amérique représentent 60 % des données disponibles. Pour *Quercus spp.* il me semble qu'il y a une légère perte d'information pour les données RENECOFOR : la nomenclature CPS signifie un site de chênes sessiles et pédonculés (espèces intimement liées par l'hybridation unidirectionnelle [12]) et donc une description plus précise que le simple genre *Quercus*.

```
cols <- c("Alpha_Number", "Species", "Site", "Reference")
mastree[!duplicated(mastree$ID) &
  mastree$Species == "Quercus spp." &
  mastree$Reference == "RENECOFOR_2020", cols]
Alpha_Number Species Site Reference
39715 5181 Quercus spp. CPS 67 RENECOFOR_2020
39729 5181 Quercus spp. CPS 77 RENECOFOR_2020
```

3.7 Species_code

C'EST un code de 6 caractères du nom de l'espèce généré en concaténant les trois premiers caractères du genre et de l'espèce et en curant manuellement pour les doublons. Pour les populations hybrides, le dernier caractère est remplacé par un X. « MIXSPP » indique une communauté de plusieurs espèces.

```
sPCODE <- mastree[!duplicated(mastree$ID), "Species_code"]
all(nchar(sPCODE) == 6L) # TRUE, OK
[1] TRUE
f <- function(x){
  genre <- unlist(strsplit(x, split = " "))[1]
  espèce <- unlist(strsplit(x, split = " "))[2]
  code <- paste0(substr(genre, 1, 3), substr(espèce, 1, 3))
  return(toupper(code))
}
check <- sapply(mastree[!duplicated(mastree$ID), "Species"], f)
ii <- which(check != sPCODE)
comp <- as.data.frame(cbind(sPCODE[ii], check[ii]))
colnames(comp) <- c("Species_code", "Check")
comp <- comp[!duplicated(comp$Species_code), ]
(comp <- comp[order(comp$Species_code), ])
Species_code Check
Acacia.craspedocarpa_x_aneura ACACRX ACACRA
Acer.pseudoplatanus ACEPSP ACEPSE
Acer.pseudosieboldianum ACEPSS ACEPSE
Carya.spp. CARSP1 CARSP1
Carpinus.spp. CARSP2 CARSP2
Eucalyptus.microcarpa EUCMCA EUCMIC
Eucalyptus.microcorys EUCMCO EUCMIC
Malus.spp. MALSP1 MALSP1
Malleastrum.spp. MALSP2 MALSP2
Mixed.species MIXSPP MIXSPP
Pinus.monophylla PINMNP PINMNP
Pinus.monticola PINMNT PINMNT
Pinus.pinaster PINPNA PINPNA
Pinus.pinea PINPNE PINPNE
Quercus.acuta QUEACA QUEACA
Quercus.acutissima QUEACI QUEACI
Quercus.mongolica QUEMNG QUEMNG
Quercus.montana QUEMNT QUEMNT
Salix.myrsinifolia_x_phylicifolia SALMYX SALMYX
```

3.8 Mono_Poly

C'EST un code d'un caractère indiquant si l'espèce est monocarpique, M, (sémelpare) ou polycarpique, P, (itéropare). La très grande majorité des espèces documentées est polycarpique :

```
all(nchar(mastree$Mono_Poly) == 1L) # TRUE, OK
[1] TRUE
table(mastree[!duplicated(mastree$ID), "Mono_Poly"])
  M   P
19 5954
```

3.9 Value

C'EST une variable numérique ou ordinale qui quantifie l'effort annuel pour la reproduction. Les valeurs sont toutes positives ou nulles :

```
any(mastree$Value < 0) # FALSE, OK
[1] FALSE
any(is.na(mastree$Value)) # FALSE, OK
[1] FALSE
```

3.10 VarType

C'EST un code d'un caractère indiquant si la série est numérique, C, ou ordinale, 0. Toutes les séries ordinales ont été recalibrées pour commencer à 1 (le plus faible effort reproductif) et ne comporter que des valeurs entières. Par exemple, une série ordinale avec les valeurs 0, 1, 2, 3 et 4 (5 modalités) est recalibrée en 1, 2, 3, 4 et 5. Il y a a peu près autant de séries ordinales que de séries quantitatives :

```
all(nchar(mastree$VarType) == 1L) # TRUE, OK
[1] TRUE
table(mastree[!duplicated(mastree$ID), "VarType"])
  C   0
2847 3126
```

3.11 Max_value

C'EST la valeur maximale de la série temporelle, tous segments confondus pour les séries comportant des lacunes. Recalculons les valeurs maximales de chaque série pour vérifier que tout va bien :

```
Max <- tapply(mastree$Max_value, mastree$ID, \(x) x[1])
checkMax <- tapply(mastree$Value, mastree$ID, max)
all.equal(Max, checkMax) # TRUE, OK
[1] TRUE
```

3.12 Unit

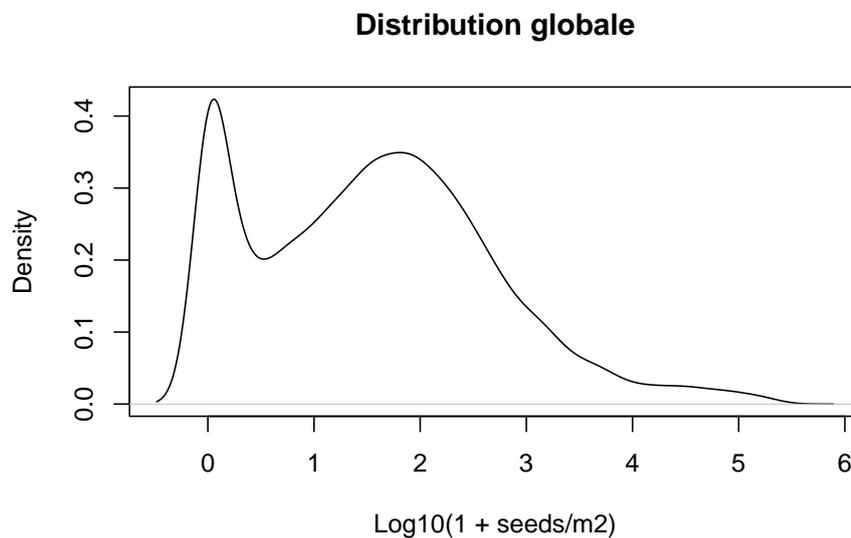
C'EST l'unité de mesure de la série temporelle. Pour les séries quantitatives elles ont été converties dans la mesure du possible pour homogénéiser les données (*e.g.* par exemple les « fruit/ha » en « fruit/m² »). La valeur `index` indique une variable quantitative sans dimension (*e.g.* une série standardisée). Les séries ordinales sont sans unité et notées NA.

```
is0 <- which(mastree$VarType == "0")
all(is.na(mastree$Unit[is0])) # TRUE, OK
```

```
[1] TRUE
head(rev(sort(table(mastree[!duplicated(mastree$ID), "Unit"],
                    useNA = "always"))))
      <NA>          seeds/m2          index
cones/individual 3126          903          325
                258          g/m2 % individuals reproducing
                244          239
```

LES unités les plus employées sont des nombres de graines par mètre carré. La distribution globale de ces valeurs est de type log-normale, enrichie en zéros, et se promène sur 6 ordres de grandeur.

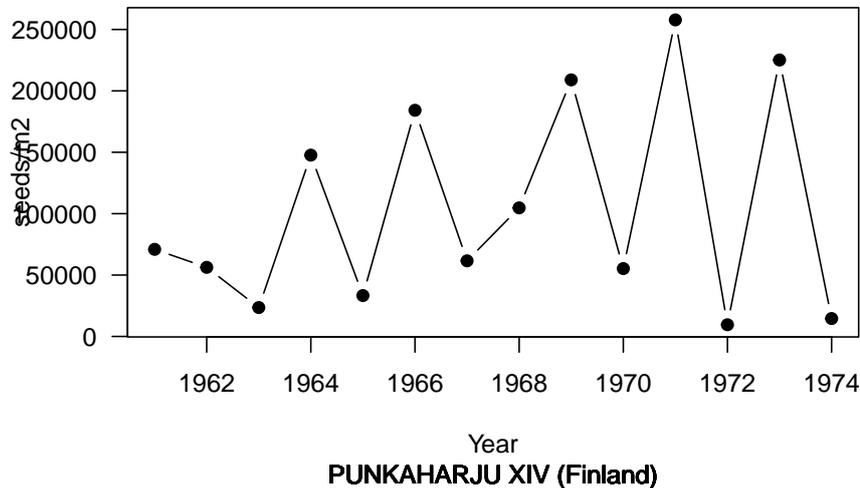
```
qui <- mastree$VarType == "C" & mastree$Unit == "seeds/m2"
x <- log10(1 + mastree[qui, "Value"])
plot(density(x), xlab = "Log10(1 + seeds/m2)",
     main = "Distribution globale")
```



LE RECORD de tous les temps a été rapporté par K. TALLQVIST [16] en 1971 pour *Betula pubescens*, le bouleau pubescent, avec plus d'un quart de million de graines par mètre carré. La masse d'une graine étant de l'ordre de 0.5 mg, cela correspond à 125 g au mètre carré.

```
plotID(with(mastree[qui, ], ID[which.max(Value)]))
```

5164-089-01-BETPUB (TALLQVIST 1978) seed



3.13 Variable

C'EST une variable qualitative qui décrit ce qui a été mesuré (cone, flower, fruit, seed, pollen, total reproduction organs).

```
rev(sort(table(mastree[!duplicated(mastree$ID), "Variable"])))
```

fruit	seed	cone
3052	1991	465
flower	pollen total reproductive organs	3
253	209	

3.14 Collection_method

C'EST une variable qualitative nominale décrivant la méthode expérimentale utilisée pour collecter les données. Pour plus de précisions il faut consulter la source originelle. De façon synthétique on a les valeurs courantes suivantes :

Seed trap : l'estimation de l'effort reproductif est basée sur le ramassage des organes reproductifs dans des collecteurs. Ces estimations peuvent être des dénombrements ou des biomasses.

Cone count, flower count, fruit count, seed count : l'estimation est basée sur le dénombrement des organes reproductifs soit directement sur la plante, soit sur le sol après leur chute.

Cone scar count, fruit scar count : l'estimation est basée sur le dénombrement de cicatrices sur les branches ou les tiges, datées sans ambiguïté, laissées par les fruits ou les cônes après leur chute.

Harvest record : l'estimation est basée sur une récolte de fruits ou de graines sur une zone de référence, souvent régionale.

Visual crop assessment : basée sur une estimation visuelle du degré d'avancement de la reproduction. Souvent sous la forme de séries ordinales, ou de séries quantitatives où l'unité est le pourcentage d'individus en phase de reproduction.

Pollen count, lake sediment pollen count : estimation basée sur le dénombrement de grains de pollen.

Dendrochronological reconstruction : estimation basée sur des méthodes de dendrochronologie.

Other quantification : autres méthodes pour estimer l'effort de reproduction. Plus de précisions peuvent être données dans « Commentaires ». Également utilisée quand les détails ne sont pas connus, en particulier pour les séries les plus anciennes (*e.g.* des séries forestières).

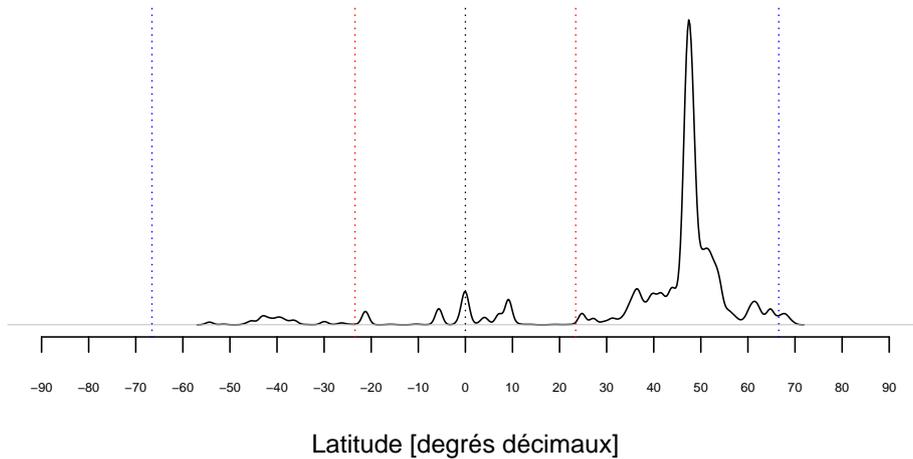
```
rev(sort(table(mastree[!duplicated(mastree$ID), "Collection_method"])))
      visual crop assessment          seed trap
      3022                        1221
      fruit count                    seed count
      361                          351
      cone count                      other quantification
      328                            270
      pollen count                    cone-scar count
      207                             102
      flower count                    harvest record
      94                              13
      lake sediment pollen count dendrochronological reconstruction
      2                                2
```

3.15 Latitude

C'EST la latitude du lieu d'observation de la série temporelle exprimée en degrés décimaux. Il y a globalement un très fort biais d'échantillonnage géographique en faveur de la zone tempérée de l'hémisphère nord :

```
main <- paste0("Distribution des latitudes\nn = ", nrow(mastree),
               " observations")
par(mar = c(5, 0, 4, 0) + 0.1)
plot(density(mastree$Latitude), xlim = c(-90, 90),
     main = main, bty = "n", yaxt = "n", ylab = "",
     xlab = "Latitude [degrés décimaux]",
     yaxt = "n")
axis(1, at = seq(-90, 90, by = 10), cex.axis = 0.5)
abline(v = c(66.55, -66.55), col = "blue", lty = 3)
abline(v = c(23.43, -23.43), col = "red", lty = 3)
abline(v = 0, lty = 3)
```

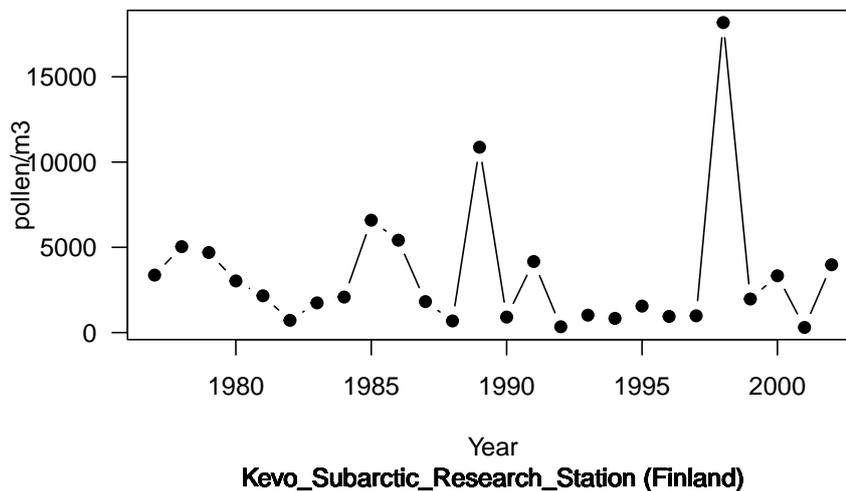
Distribution des latitudes n = 73828 observations



PAR curiosité on cherche quelle est la série la plus septentrionale. C'est une étude [11] du suivi de la concentration pollinique de *Betula pubescens* à la station sub-artique de Kevo en Finlande.

```
plotID(with(mastree[!duplicated(mastree$ID), ], ID[which.max(Latitude)]))
```

2512-001-01-BETPUB (Klemola_2003) pollen

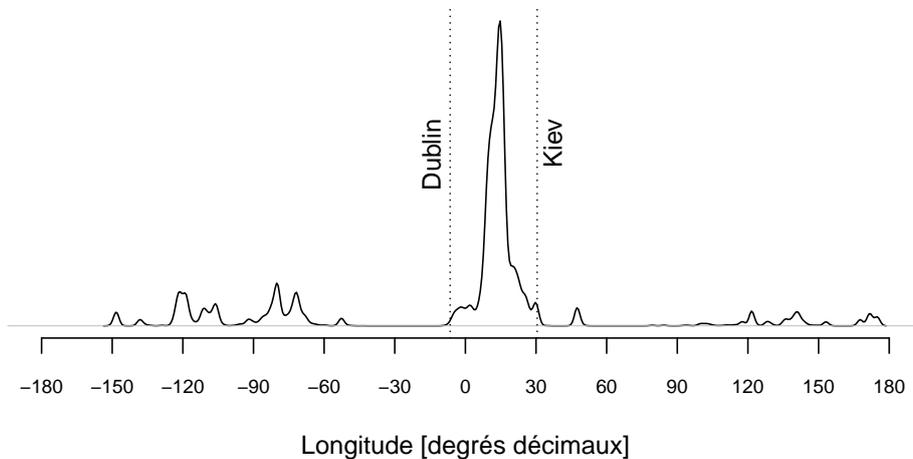


3.16 Longitude

C'EST la longitude du lieu d'observation de la série temporelle exprimée en degrés décimaux. Il y a un très fort biais d'échantillonnage géographique en faveur des longitudes européennes :

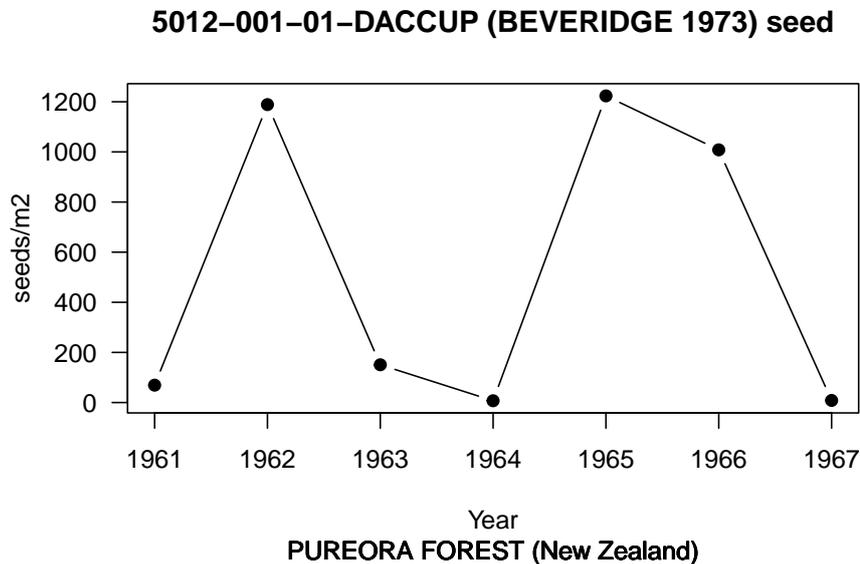
```
main <- paste0("Distribution des longitudes\nn = ", nrow(mastree),
              " observations")
par(mar = c(5, 0, 4, 0) + 0.1)
plot(density(mastree$Longitude), xlim = c(-180, 180),
     main = main, bty = "n", yaxt = "n", ylab = "",
     xlab = "Longitude [degrés décimaux]",
     xaxt = "n")
axis(1, at = seq(-180, 180, by = 30), cex.axis = 0.75)
abline(v = -6.5, lty = 3)
text(-6.5, 0.05, "Dublin", srt = 90, pos = 2, xpd = NA)
abline(v = 30.5, lty = 3)
text(45, 0.05, "Kiev", srt = 90, pos = 2, xpd = NA)
```

Distribution des longitudes n = 73828 observations



PAR curiosité on cherche quelle est la série la plus aux antipodes de la méridienne de l'origine. C'est une étude [5] sur *Dacrydium cupressinum*, *Dacrycarpus dacrydioides*, *Podocarpus totara*, *Prumnopitys ferruginea* et *P. taxifolia* menée dans la forêt de Pureora en Nouvelle-Zélande.

```
plotID(with(mastree[!duplicated(mastree$ID), ], ID[which.max(Longitude)]))
```



3.17 Coordinate_flag

C'EST une variable qualitative ordonnée codée sur un caractère indiquant la précision des données de géolocalisation, avec les valeurs A, B et C dans l'ordre décroissant :

A : coordonnées fournies dans la source originelle.

B : coordonnées estimées par le compilateur sur la base d'une carte ou d'autres informations de géolocalisation fournies par la source originelle.

C : coordonnées estimées par le compilateur comme le centre approximatif de la plus petite entité géographique fournie dans la source originelle (*e.g.* division administrative, état, île) ou estimée à partir de la description du site (*e.g.* localisation par rapport à un habitat ou un point géographique caractéristique).

```
all(nchar(mastree$Coordinate_flag) == 1L) # TRUE, OK
[1] TRUE
with(mastree[!duplicated(mastree$ID), ], table(Coordinate_flag))
Coordinate_flag
  A   B   C
4195 915 863
```

3.18 Site

C'EST une variable qualitative nominale donnant en clair le nom du site de l'étude d'après les informations de la source originelle. On peut utiliser `Site_number` pour créer des clefs d'identification des séries, mais `Site` permet de faire plus facilement le lien avec la source originelle.

```
class(mastree$Site)
[1] "character"
tmp <- unique(mastree$Site)
table(Encoding(tmp))
unknown   UTF-8
 2499     18

isutf8 <- Encoding(tmp) == "UTF-8"
tmp[isutf8]

[1] "Hawkin's Hill"
[2] "University of South Florida's Ecological Research Area"
[3] "N'washitsumbe enclosure"
[4] "Värriö, Site 1"
[5] "Värriö, Site 2"
[6] "Värriö, Site 3"
[7] "Värriö, Site 4"
[8] "Schönbuch nature reserve"
[9] "Baden-W\x8irttemberg"
[10] "Huerto Ram\xairez"
[11] "San Bartolom\x82"
[12] "Villaviciosa de Córdoba, Cabeza Aguda"
[13] "Villaviciosa de Córdoba, Cañadas del Névalo"
[14] "Villaviciosa de Córdoba, Las Parrillas"
[15] "Chojnów Forest District"
[16] "Caracara\xed - ITA"
[17] "S\xee3o Jo\xee3o da Baliza"
[18] "PU\xc9CHABON STATE FOREST"

Encoding(tmp) <- "latin1"
tmp[isutf8]

[1] "Hawkinâ€s Hill"
[2] "University of South Floridaâ€s Ecological Research Area"
[3] "Nâ€washitsumbe enclosure"
[4] "VÃ=rrriÃ€, Site 1"
[5] "VÃ=rrriÃ€, Site 2"
[6] "VÃ=rrriÃ€, Site 3"
[7] "VÃ=rrriÃ€, Site 4"
[8] "SchÃ€nbuch nature reserve"
[9] "Baden-W<81>rttemberg"
[10] "Huerto Ram;rez"
[11] "San Bartolom,"
[12] "Villaviciosa de CÃ³rdoba, Cabeza Aguda"
[13] "Villaviciosa de CÃ³rdoba, CaÃ±adas del NÃ©valo"
[14] "Villaviciosa de CÃ³rdoba, Las Parrillas"
[15] "ChojnÃ³w Forest District"
[16] "Caracarái - ITA"
[17] "São João da Baliza"
[18] "PUÉCHABON STATE FOREST"
```

L'ENCODAGE des caractères n'est presque jamais renseigné pour cette variable. Même quand il est déclaré en UTF-8 c'est parfois correct avec l'apostrophe typographique de « Hawkin's », « Florida's » et « N'washitsumbe » et les caractères diacrités de « Värriö », « Schönbuch », « Córdoba », « Cañadas del Névalo » et « Chojnów ». Mais la présence de mojibakes montre que c'est parfois incorrect, on serait plutôt en présence de latin1 pour « Caracarái », « São João » et « PUÉCHABON » et d'un autre encodage pour les items 9 à 11 ci-dessus.

3.19 Country

C'EST le nom du pays où ont été faites les observations. C'est la version courte en anglais (ISO3166-1) standardisée avec le paquet  countrycode [1]. On retrouve ici le biais d'échantillonnage en faveur des pays de l'hémisphère Nord.

```
head(rev(sort(table(mastree[!duplicated(mastree$ID), "Country"]))), n = 10)
```

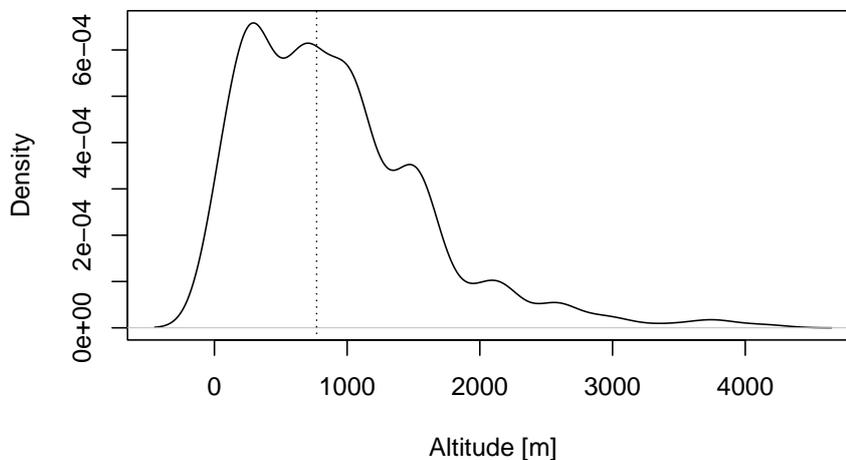
Austria	2231	United States of America (the)	750
Germany	327	Canada	193
Finland	181	Poland	163
Italy	155	Spain	145
Japan	141	New Zealand	126

3.20 Elevation

C'EST l'altitude en mètres au dessus du niveau de la mer telle qu'indiquée dans la source originelle. Codé NA si l'information est manquante. Ce sont majoritairement des données de plaine puisque plus de la moitié des séries sont en dessous de 1000 mètres d'altitude :

```
with(mastree[!duplicated(mastree$ID), ], {
  x <- Elevation[!is.na(Elevation)]
  main <- paste("Distribution des altitudes\nn =", length(x), "séries")
  plot(density(x),
        main = main,
        xlab = "Altitude [m]")
  abline(v = median(x), lty = 3)
})
```

**Distribution des altitudes
n = 1196 séries**

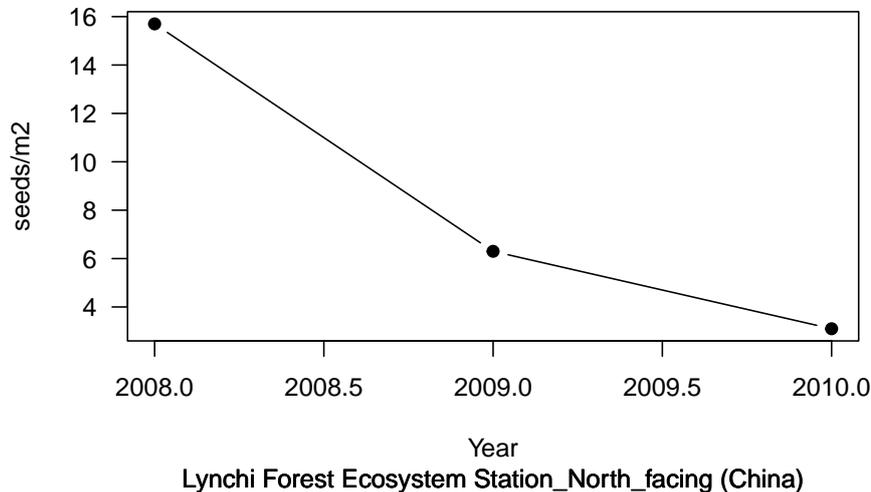


PAR curiosité on cherche quelle est la série à la plus haute altitude. C'est une étude¹⁶ sur les graines de *Abies forrestii* à 4200 mètres d'altitude dans la station de l'écosystème de la forêt de Lynchin en Chine.

```
plotID(with(mastree[!duplicated(mastree$ID), ], ID[which.max(Elevation)]))
```

¹⁶La référence Alpha_number = 2280 pointe vers « Wells, K., and R. Bagchi. 2005. Eat in or take away - Seed predation and removal by rats (muridae) during a fruiting event in a dipterocarp rainforest. Raffles Bulletin of Zoology 53:281-286 » mais cela ne correspond pas à Reference = Wang_2016 pour cette série.

2280-004-01-ABIFOR (Wang_2016) seed



3.21 Spatial_unit

C'EST une variable qualitative ordonnée indiquant l'étendue spatiale de la série telle que décrite dans la source originelle. Elle possède 4 modalités qui dans l'ordre croissant sont :

stand : $\leq 10^2$ ha

patch : $10^2 - 10^4$ ha

region : $10^4 - 10^6$ ha

super-region : $> 10^6$ ha

Il y a un biais d'échantillonnage en faveur des séries portant sur des petites étendues spatiales :

```
mastree$Spatial_unit <- factor(mastree$Spatial_unit,
  levels = c("stand", "patch", "region", "super-region"),
  ordered = TRUE)
with(mastree[!duplicated(mastree$ID), ], table(Spatial_unit))
```

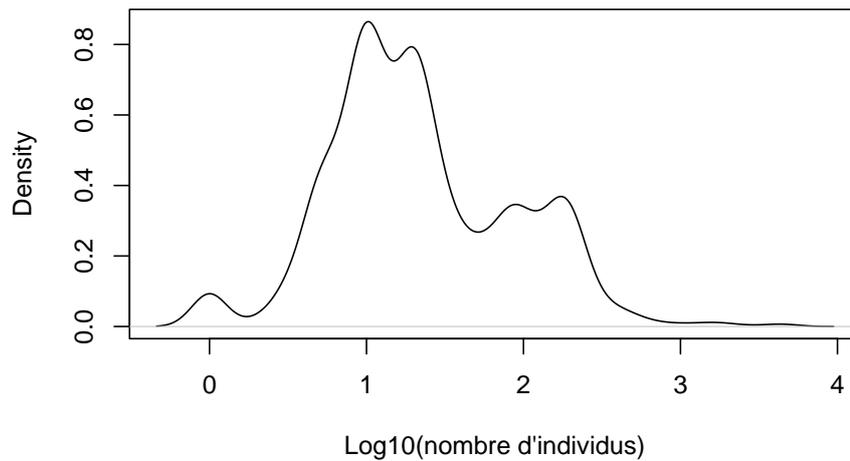
Spatial_unit	stand	patch	region	super-region
	2024	762	428	520

3.22 No_indivs

C'EST le nombre de végétaux suivis ou le nombre de collecteurs. La valeur 9999 indique que même si le nombre d'individus n'est pas indiqué, le compilateur a estimé que cette valeur est très certainement supérieure ou égale à 10. La valeur NA indique qu'il n'y avait pas d'information sur la taille de l'échantillon dans la source originelle. Les valeurs sont très variables d'une série à l'autre et courent sur 4 ordres de grandeur, avec un pic au voisinage de 10 individus suivis :

```
with(mastree[!duplicated(mastree$ID), ], {
  x <- No_indivs[!is.na(No_indivs) & No_indivs != 9999]
  main <- paste("Distribution de la taille des échantillons\nn =",
    length(x), "séries")
  plot(density(log10(x)), main = main,
    xlab = "Log10(nombre d'individus)")
})
```

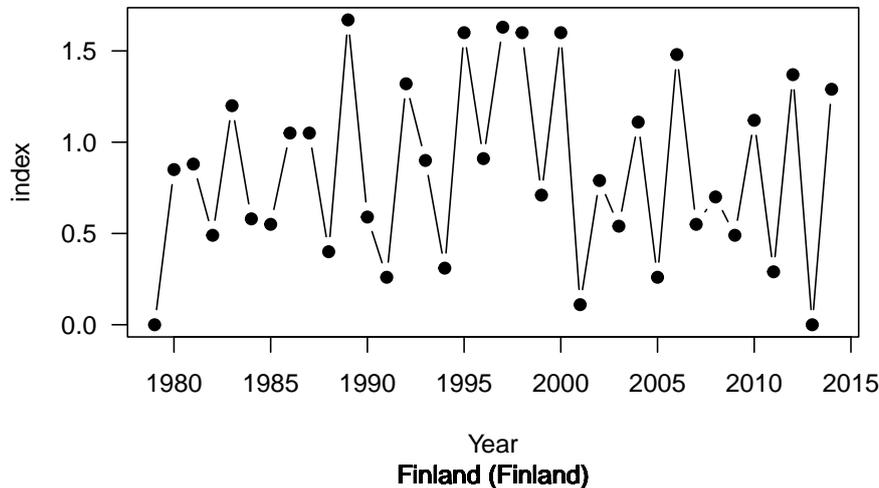
**Distribution de la taille des échantillons
n = 2185 séries**



PAR curiosité on cherche la série portant sur le plus grand nombre d'individus.
C'est un suivi [17] du nombre de cônes chez *Picea abies* en Finlande.

```
with(mastree[!duplicated(mastree$ID), ], {
  qui <- !is.na(No_indivs) & No_indivs != 9999
  x <- No_indivs[qui]
  names(x) <- ID[qui]
  plotID(names(x[which.max(x)]))
})
```

3054-001-01-PICABI (Gallego Zamorano et al. 2016) cone



3.23 Start

C'est la première année de la série tous segments confondus.

```
Starts <- tapply(mastree$Start, mastree$ID, \(x) x[1])
checkStart <- tapply(mastree$Year, mastree$ID, min)
all.equal(Starts, checkStart) # TRUE, OK
[1] TRUE
```

3.24 End

C'est la dernière année de la série tous segments confondus.

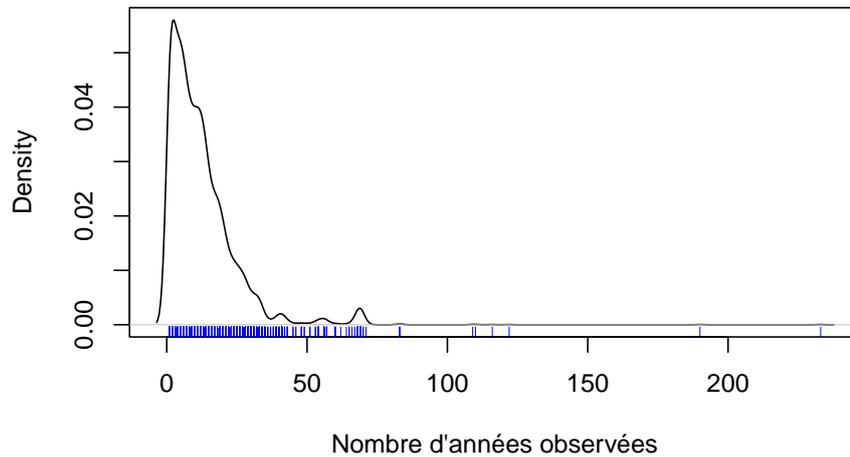
```
Ends <- tapply(mastree$End, mastree$ID, \(x) x[1])
checkEnd <- tapply(mastree$Year, mastree$ID, max)
all.equal(Ends, checkEnd) # TRUE, OK
[1] TRUE
```

3.25 Length

C'EST le nombre d'années documentées dans la série. Ce n'est pas forcément le nombre d'années entre le début et la fin de la série puisqu'elle peut présenter des lacunes.

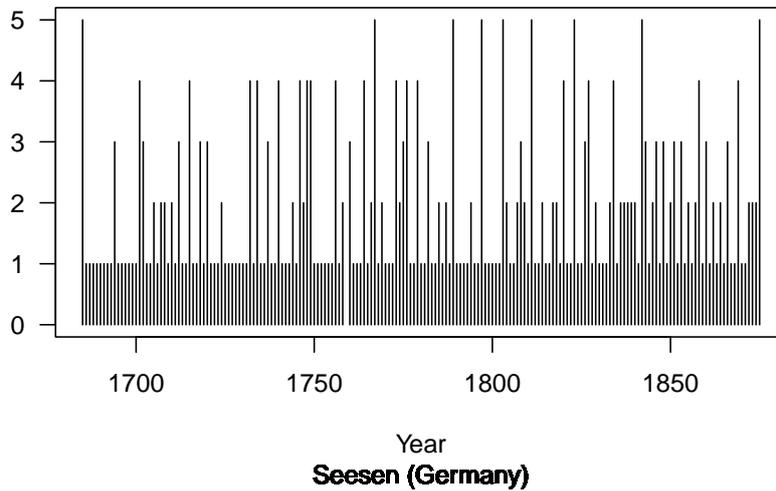
```
Lengths <- tapply(mastree$Length, mastree$ID, \(x) x[1])
checkLength <- tapply(mastree$Year, mastree$ID, length)
all.equal(Lengths, checkLength) # TRUE, OK
[1] TRUE
with(mastree[!duplicated(mastree$ID), ], {
  x <- Length
  main <- paste("Distribution de la longueur des séries\nn =",
               length(x), "séries")
  plot(density(x), main = main,
       xlab = "Nombre d'années observées")
  rug(x, col = "blue")
})
```

Distribution de la longueur des séries n = 5973 séries



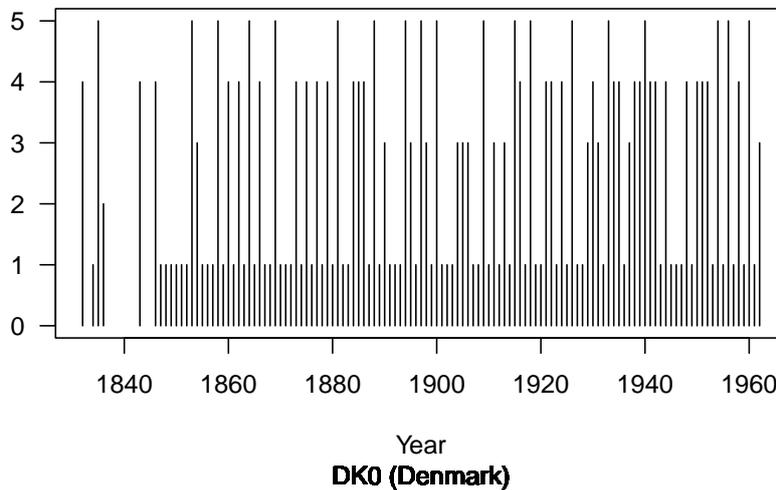
PAR curiosité on cherche quelles sont les séries ayant plus de 100 années observées. La plus longue correspond à la série la plus ancienne (voir section 3.5 page 13). La suivante est une étude [4] de 1877 de suivi du nombre de graines chez *Fagus sylvatica* à Seesen en Allemagne :

```
(maxlts <- with(mastree[!duplicated(mastree$ID), ], {
  qui <- Length > 100
  x <- Length[qui]
  names(x) <- ID[qui]
  rev(sort(x))
}))
3067-001-01-FAGSYL 3014-001-01-FAGSYL 3067-006-01-FAGSYL 3125-001-01-FAGSYL
                233                190                122                116
3079-001-01-FAGSYL 3025-001-01-FAGSYL
                110                109
plotID(names(maxlts)[2], type = "h", ylim = c(0, 5))
```

3014-001-01-FAGSYL (Beling 1877) seed

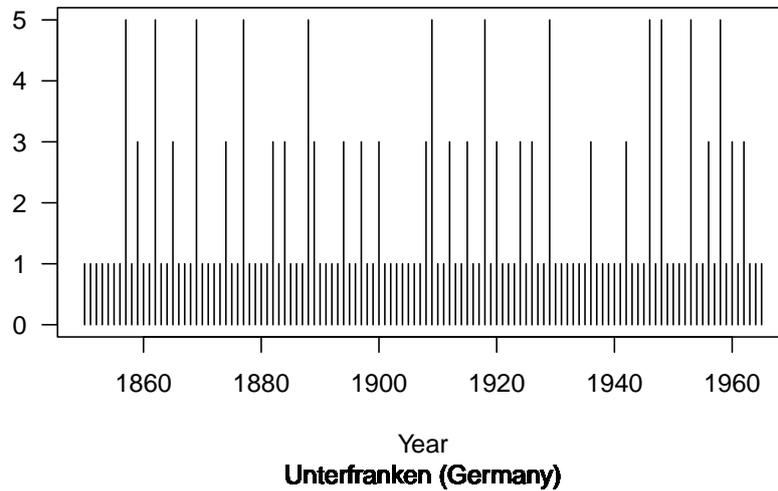
L 'ÉTUDE suivante [9] de 1964 est également un suivi des graines de *Fagus sylvatica* sur le site DK0 au Danemark.

```
plotID(names(maxlts)[3], type = "h", ylim = c(0, 5))
```

3067-006-01-FAGSYL (Hase 1964) seed

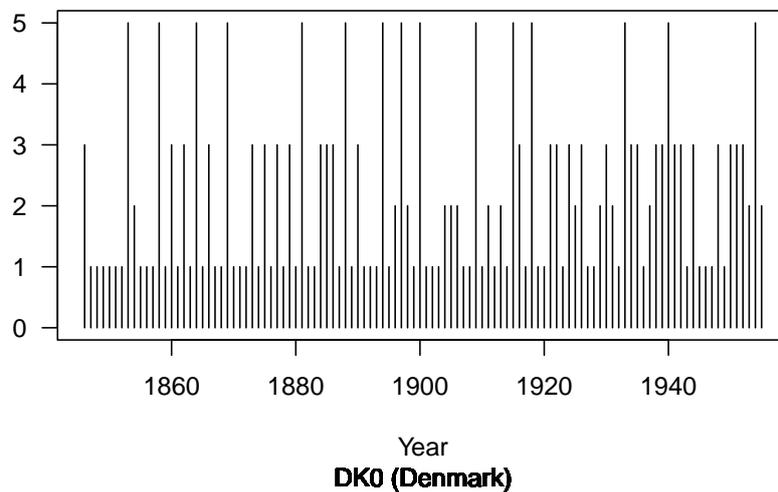
L 'ÉTUDE suivante [13] de 1964 est également un suivi des graines de *Fagus sylvatica* à Unterfranken en Allemagne.

```
plotID(names(maxlts)[4], type = "h", ylim = c(0, 5))
```

3125-001-01-FAGSYL (Maurer 1964) seed

L 'ÉTUDE suivante [10] de 1960 est également un suivi des graines de *Fagus sylvatica* sur le site DKO au Danemark.

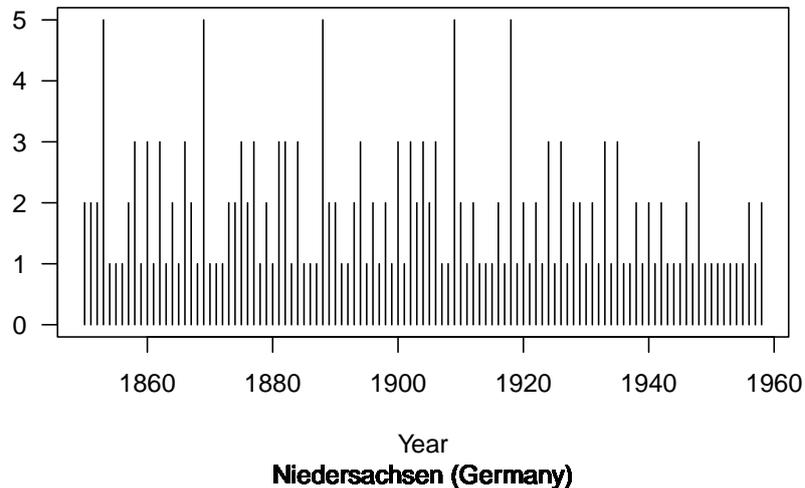
```
plotID(names(maxlts)[5], type = "h", ylim = c(0, 5))
```

3079-001-01-FAGSYL (Holmsgaard & Olsen 1960) seed

L 'ÉTUDE suivante [6] de 1958 est également un suivi des graines de *Fagus sylvatica* à Niedersachsen en Allemagne.

```
plotID(names(maxlts)[6], type = "h", ylim = c(0, 5))
```

3025-001-01-FAGSYL (Borchers 1958) seed



3.26 Reference

```
tail(unique(mastree$Reference), n = 10)
[1] "DeRose_unpublished" "Redmond_2013" "Shibata_Mitsure_2019"
[4] "Hertel_2017" "Auger_2017" "Dunham_2018"
[7] "Chang-Yang_2015" "Bush_2020" "Kalbitzer_2021"
[10] "Gratzer_2021"
```

C'EST un codage de la source originelle sous une forme plus mnémotechnique que le pointeur `Alpha_Number`, du type auteur/année. La liste complète des références est donnée dans l'annexe 4 de [8]. Le premier auteur de [8] conserve le PDF des sources originelles.

3.27 Record_type

C'EST une variable qualitative nominale précisant la nature de la source des données originelle. Elle possède trois modalités :

Peer-reviewed : données issues de revues scientifiques classiques avec comité de lecture composé de spécialistes du domaine

Grey : données issues de la littérature grise (pas d'évaluation par un comité d'expert, le présent document est un exemple de littérature grise).

unpublished : données non déjà publiées auparavant.

```
mastree$Record_type <- as.factor(mastree$Record_type)
with(mastree[!duplicated(mastree$ID), ], table(Record_type))
Record_type
  grey peer-reviewed Tissier_2020 unpublished Unpublished unpublished
    443           2116             6          3381             21             6

# Correction typos
levels(mastree$Record_type)[3:6] <- c("peer-reviewed", rep("unpublished", 3))
with(mastree[!duplicated(mastree$ID), ], table(Record_type))
```

```
Record_type
  grey peer-reviewed unpublished
   443   2122         3408
```

ON voit ici tout l'intérêt de MASTREE+ : utiliser très facilement des données qui sont habituellement difficilement accessibles car soit non publiées, ou publiées de façon confidentielle (et dans une langue que l'on ne maîtrise pas forcément).

3.28 ID_enterer

```
with(mastree[!duplicated(mastree$ID), ], table(ID_enterer))
ID_enterer
  AHP  DA  DV  ES  IP  JL  JVD  JVM  TC  XTT
3034 983 282 49 709 152 398 176 107 83
```

C'EST l'identifiant du compilateur responsable de l'extraction et de l'incorporation des données dans MASTREE+ :

AHP : Andrew HACKET-PAIN ;

DA : Davide ASCOLI ;

DV : Davide VECCHIO ;

ES : Éliane SCHERMER ;

IP : Ian PEARSE ;

JL : Jalene LAMONTAGNE ;

JVD : Joep VAN DORMOLEN ;

JVM : Jose MORIS ;

TC : Thomas CAIGNARD ;

XTT : Tingting XUE.

3.29 Date_entry

C'EST la date d'incorporation de la série dans MASTREE+ au format yyyy-mm-dd, c'est à dire le format ISO 8601. Comme c'est le format par défaut pour `as.Date()` ce sera assez simple pour les manipuler.

```
mastree$Date_entry <- as.Date(mastree$Date_entry)
with(mastree[!duplicated(mastree$ID), ], summary(Date_entry))
      Min.      1st Qu.      Median      Mean      3rd Qu.      Max.
"2019-02-26" "2020-07-03" "2020-07-21" "2020-08-20" "2020-09-03" "2022-01-18"
```

3.30 Note on data location

Ce sont des notes prises par le compilateur sur la localisation des données dans la publication originelle, telles que le numéro de page, de figure ou de table. Les valeurs non renseignées sont codées NA.

```
head(unique(mastree$`Note on data location`), n = 10)
[1] "p. 403"
[2] NA
[3] "approximative extraction from the map in Fig. 1"
[4] "Table 3, converted to /m2"
[5] "Table 2"
[6] "Figure 4"
[7] "Figure 2"
[8] "text, p. 12"
[9] "Figure 4a"
[10] "Figure 5a"
```

3.31 Comments

Ce sont des commentaires du compilateur, des précisions fournies par les contributeurs des données non publiées ou des notes d'Andrew HACKETT-PAIN ou Jessie J. FOEST pendant la phase de validation des données.

3.32 Study_ID

La colonne Study_ID permet de tracer la source originelle des données :

- 1° M* données extraites de la littérature, voir la référence correspondante dans l'annexe S4 de [8].
- 2° A* données issues des compilations d'ASCOLI et collaborateurs [2, 3].
- 3° P* données issues de la compilation de PEARSE et collaborateurs [14].
- 4° D* données non publiées avant [8].

```
rev(sort(table(substring(mastree$Study_ID, 1 , 1))))
      D      A      P      M
34994 17364 11388 10082
```

3.33 Sauvegarde

On sauvegarde ici le jeu de données après correction de quelques typos.

```
save(mastree, file = "mastree.Rda")
```

References

- [1] V. Arel-Bundock, N. Enevoldsen, and C.J. Yetman. countrycode: An R package to convert country names and country codes. *Journal of Open Source Software*, 3(28):848, 2018.
- [2] D. Ascoli, A. Hacket-Pain, J.M. LaMontagne, A. Cardil, M. Conedera, J. Maringer, R. Motta, I.S. Pearse, and G. Vacchiano. Climate teleconnections synchronize *Picea glauca* masting and fire disturbance: Evidence for a fire-related form of environmental prediction. *Journal of Ecology*, 108:1186–1198, 2020.
- [3] D. Ascoli, J. Maringer, A. Hacket-Pain, M. Conedera, I. Drobyshev, R. Motta, M. Cirolli, W. Kantorowicz, C. Zang, S. Schueler, L. Croisé, P. Piussi, R. Berretti, C. Palaghianu, M. Westergren, J.G.A. Lageard, A. Burkart, R.G. Bichsel, P.A. Thomas, B. Beudert, R. Övergaard, and G. Vacchiano. Two centuries of masting data for european beech and norway spruce across the european continent. *Ecology*, 98:1473–1473, 2017.
- [4] K.W.T. Beling. Über die Samenjahre der Eiche, Buche und Fichte. *Monatschrift für das Forst-und Jagdwesen*, 21:49–81, 1877.
- [5] A.E. Beveridge. Regeneration of podocarps in a central north island forest. *New Zealand journal of forestry*, 1973.
- [6] K. Borchers. Auswirkungen rezenter Klimaschwankungen auf die Häufigkeit von Buchen-Samenjahren in Niedersachsen. *Forst-und Holzwirtschaft*, 13:330–333, 1958.
- [7] L. Cayuela, I. Macarro, A. Stein, and J. Oksanen. *Taxonstand: Taxonomic Standardization of Plant Species Names*, 2021. R package version 2.4.
- [8] A. Hacket-Pain, J.J. Foest, Ian S. Pearse, Jalene M. LaMontagne, Walter D. Koenig, Giorgio Vacchiano, Michał Bogdziewicz, Thomas Caignard, Paulina Celebias, Joep van Dormolen, Marcos Fernández-Martínez, Jose V. Moris, Ciprian Palaghianu, Mario Pesendorfer, Akiko Satake, Eliane Schermer, Andrew J. Tanentzap, Peter A. Thomas, Davide Vecchio, Andreas P. Wion, Thomas Wohlgemuth, Tingting Xue, Katharine Abernethy, Marie-Claire Aravena Acuña, Marcelo Daniel Barrera, Jessica H. Barton, Stan Boutin, Emma R. Bush, Sergio Donoso Calderón, Felipe S. Carevic, Carolina Volkmer de Castilho, Juan Manuel Cellini, Colin A. Chapman, Hazel Chapman, Francesco Chianucci, Patricia da Costa, Luc Croisé, Andrea Cutini, Ben Dantzer, R. Justin DeRose, Jean-Thoussaint Dikangadissi, Edmond Dimoto, Fernanda Lopes da Fonseca, Leonardo Gallo, Georg Gratzer, David F. Greene, Martín A. Hadad, Alejandro Huertas Herrera, Kathryn J. Jeffery, Jill F. Johnstone, Urs Kalbitzer, Władysław Kantorowicz, Christie A. Klimas, Jonathan G. A. Lageard, Jeffrey Lane, Katharina Lapin, Mateusz Ledwoń, Abigail C. Leeper, Maria Vanessa Lencinas, Ana Cláudia Lira-Guedes, Michael C. Lordon, Paula Marchelli, Shealyn Marino, Harald Schmidt Van Marle, Andrew G. McAdam, Ludovic R. W. Momont, Manuel Nicolas, Lúcia Helena de Oliveira Wadt, Parisa Panahi, Guillermo Martínez Pastur, Thomas Patterson, Pablo Luis Peri, Łukasz

- Piechnik, Mehdi Pourhashemi, Claudia Espinoza Quezada, Fidel A. Roig, Karen Peña Rojas, Yamina Micaela Rosas, Silvio Schueler, Barbara Seget, Rosina Soler, Michael A. Steele, Mónica Toro-Manríquez, Caroline E. G. Tutin, Tharcisse Ukizintambara, Lee White, Biplang Yadok, John L. Willis, Anita Zolles, Magdalena Żywiec, and Davide Ascoli. Mastree+: time-series of plant reproductive effort from six continents. *Global Change Biology*, 00:1–17, 2022.
- [9] W. Hase. Die Buchenmast in Schleswig-Holstein und ihre Abhängigkeit von der Witterung. *Mitt. Deutsch. Wetterdienst*, 31:3–45, 1964.
- [10] E. Holmsgaard and H.C. Olsen. The influence of weather on beech mast. *Forstlige Forsøgsvaesen I Danmark*, 26:347–370, 1960.
- [11] T. Klemola, S. Hanhimäki, K. Ruohomäki, J. Senn, M. Tanhuanpää, P. Kaitaniemi, H. Ranta, and E. Haukioja. Performance of the cyclic autumnal moth, *Epirrita autumnata*, in relation to birch mast seeding. *Oecologia*, 135:354–361, 2003.
- [12] A. Kremer, R. Petit, and A. Ducousso. Biologie évolutive et diversité génétique des chênes sessile et pédonculé. *Revue forestière française*, 54:111–130, 2002.
- [13] E. Maurer. Buchen-und Eichensamenjahre in Unterfranken während der letzten 100 Jahre. *Allgemeine Forstzeitschrift*, 31:469–470, 1964.
- [14] I.S. Pearse, J.M. LaMontagne, and W.D. Koenig. Inter-annual variation in seed production has increased over time (1900-2014). *Proceedings of the Royal Society B: Biological Sciences*, 284(1868):20171666, 2017.
- [15] Sun Microsystems. XDR: external data representation standard. RFC 1014. Technical report, Network Working Group, 1987.
- [16] K. Tallqvist. Results of long time measurements of the quality of flowering and seed crop of trees. *Folia Forestali*, 364:1–60, 1978.
- [17] J.G. Zamorano, T. Hokkanen, and A. Lehikoinen. Climate-driven synchrony in seed production of masting deciduous and conifer tree species. *Journal of Plant Ecology*, 11:180–188, 2018.