

Géolocalisation de 11 espèces de chênes observées en France métropolitaine entre 2005 et 2021

Pr Jean R. LOBRY

Où sont les espèces chênes en France métropolitaine de nos jours ?
Les données de l'inventaire forestier national français donnent la géolocalisation de relevés de l'abondance de 11 espèces d'iceux entre 2005 et 2021 sur des placettes boisées et observables de 7 ares aléatoirement posées sur un maillage fin au kilomètre carré du territoire. Des données sont disponibles pour environ les deux tiers des mailles possibles.

Table des matières

1	Prolégomène à l'utilisation des données	2
1.1	Source et documentation	2
1.2	Les placettes PLACETTE.csv	6
1.3	La flore FLORE.CSV	14
2	Localisation des chênes	16
2.1	Principe des cartes	16
2.2	Les rois de la forêt	18
2.3	Les outsiders	21
2.4	Les méridionaux	24
2.5	Espèces introduites récemment	25
	Références	29

1 Prolégomène à l'utilisation des données

LES données de l'inventaire forestier national français sont très riches et représentent une somme de travail colossale. La documentation [7] est très claire sur le fait que « les données brutes *ne peuvent absolument pas permettre de recalculer des résultats d'inventaire* car l'utilisateur ne dispose ni du poids des points, ni des règles de post-stratification utilisées par le service de calcul de l'inventaire forestier. » Nous ne ferons ici qu'effleurer la surface en exploitant les données *observationnelles* sur l'abondance des chênes. Ces données ne représentent pas un instantané de la situation mais la compilation de 17 campagnes annuelles s'échelonnant de 2005 à 2021¹. Or, cette situation est loin d'être statique puisque, par exemple, la surface boisée en France métropolitaine a augmenté de l'ordre de 1 % par an ces 30 dernières années². Cependant, dans le cas qui nous intéresse, cette temporalité est faible par rapport à l'espérance de vie d'un chêne qui est de l'ordre de 5 siècles³. Ainsi, même si nous ne disposons pas d'une photo instantanée de la situation, le temps d'exposition est relativement faible par rapport à la vitesse d'évolution du sujet pour qu'il soit raisonnable d'espérer de ne pas avoir trop de « flou de bougé » dans le rendu final.

L'OBJECTIF de cette introduction est d'expliquer de façon *reproductible* comment nous nous sommes approprié les données sous . Le lecteur pressé pourra aller directement à la section 2 page 16 pour accéder aux cartes, au risque de passer à coté d'erreurs potentiellement commises lors de cette appropriation.

1.1 Source et documentation

LES données sont celles de l'IGN⁴, après décompression le dossier contient les données et la documentation afférente : un document d'introduction général [7] et des documents plus spécifiques relatifs aux fichiers de données⁵. Le contenu du dossier décompressé est le suivant :

```
dir("export_dataifn_2005_2021/")
[1] "ARBRE.csv"                "BOIS_MORT.csv"
[3] "COUVERT.csv"             "ECOLOGIE.csv"
[5] "espar-cdref13.csv"       "FLORE.csv"
[7] "HABITAT.csv"             "IGN_DB_doc_arbre.pdf"
[9] "IGN_DB_doc_bois_mort.pdf" "IGN_DB_doc_couvert.pdf"
[11] "IGN_DB_doc_ecologie.pdf" "IGN_DB_doc_flore.pdf"
[13] "IGN_DB_doc_habitat.pdf"  "IGN_DB_doc_placette.pdf"
[15] "IGN_DB_documentation_generale.pdf" "metadonnees.csv"
[17] "PLACETTE.csv"
```

EN plus des 7 fichiers *.csv de données et des 8 fichiers de documentation associés *.pdf on trouve les fichiers :

¹La campagne t commence en octobre de l'année $t - 1$ et se termine en octobre de l'année t .

²<https://inventaire-forestier.ign.fr/spip.php?rubrique11>

³En cas de sylviculture, la durée de renouvellement d'une futaie de chênes est de l'ordre du siècle et pour une exploitation en taillis de l'ordre du demi-siècle.

⁴IGN Inventaire forestier national français, Données brutes, Campagnes annuelles 2005 et suivantes, <https://inventaire-forestier.ign.fr/dataIFN/>, site consulté le 18/04/2023. Une sauvegarde de la version utilisée ici est disponible à <http://pbil.univ-lyon1.fr/R/donnees/dataIFN/>

⁵ARBRE.csv [9], BOIS_MORT.csv [1], COUVERT.csv [2], ECOLOGIE.csv [3], FLORE.csv [4], HABITAT.csv [5] et PLACETTE.csv [6].

1. `espar-cdref13.csv`;
2. `metadonnees.csv`.

1.1.1 `espar-cdref13.csv`

CE fichier permet de faire la correspondance entre les clefs d'identification des noms vernaculaires (variable `ESPAR`⁶ pour « espèce arborée ») et scientifiques (variable `CD_REF`⁷ pour « taxon de référence » selon la version 13 de la nomenclature du MNHN [12]⁸) des espèces :

```
esparcdref <- read.table("export_dataifn_2005_2021/espar-cdref13.csv",
  sep = ";", header = TRUE, quote = "")
colnames(esparcdref)[1] <- "espar" # à cause du // en ligne 1
any(duplicated(esparcdref$espar)) # Sanity check
[1] FALSE
any(duplicated(esparcdref$cd_ref)) # Sanity check
[1] TRUE
subset(esparcdref, cd_ref == esparcdref[duplicated(esparcdref$cd_ref), "cd_ref"])
  espar      lib_espar cd_ref      lib_cdref
138 49TG Tamaris de France 125426 Tamarix gallica
193 49TA Tamaris de France 125426 Tamarix gallica
```

ON peut voir que la variable `ESPAR` n'est pas une clef pour l'identification taxonomique puisque 49TG et 49TA correspondent tous deux à *Tamarix gallica*, mais ce n'est pas très grave puisque seuls les chênes nous intéressent ici. Voyons les espèces de chênes représentées, on recherche la chaîne de caractères « Chênes » dans les noms vernaculaires, puis la chaîne de caractères « Quercus » dans les noms scientifiques. On vérifie que l'on obtient bien le même résultat dans les deux cas.

```
gch <- esparcdref[grepl("Chêne", esparcdref$lib_espar), ]
gqu <- esparcdref[grepl("Quercus", esparcdref$lib_cdref), ]
identical(gch, gqu) # Sanity check
[1] TRUE
gch
  espar      lib_espar cd_ref      lib_cdref
98   34   Chêne chevelu 116670   Quercus cerris
99  29CE   Chêne écarlate 159897   Quercus coccinea
100   6     Chêne vert 116704   Quercus ilex
101   5   Chêne pubescent 116751   Quercus pubescens
102 29CM   Chêne des marais 116740   Quercus palustris
103   2   Chêne pédonculé 139596   Quercus robur var. robur
104  08P   Chêne faux-liège 116677   Quercus crenata
105   4     Chêne rouge 116762   Quercus rubra
106   3   Chêne sessile 139584   Quercus petraea subsp. petraea
107  08S     Chêne-liège 116774   Quercus suber
108   7     Chêne tauzin 116754   Quercus pyrenaica
```

Il y a donc 11 espèces de chênes documentées dans le fichier `espar-cdref13.csv`.

⁶Utilisée dans `ARBRE.csv`, `BOIS_MORT.csv` et `COUVERT.csv`

⁷Utilisée dans `FLORE.csv`

⁸<https://inpn.mnhn.fr/telechargement/referentielEspece/referentielTaxo>

1.1.2 metadonnees.csv

LES variables et les unités peuvent changer d'une campagne à l'autre. Le fichier `metadonnees.csv` permet de connaître les campagnes pour lesquelles une donnée et son unité sont valides, par exemple pour la variable `ESPAR` :

```
meta <- readLines("export_dataifn_2005_2021/metadonnees.csv")
meta[substr(meta, 1, 11) == "ESPAR;ESPAR"]
[1] "ESPAR;ESPAR;2005 - 2006;Qualitatif;Espèce arborée;Espèce botanique d'arbre ou groupe d'espèces détaillé;"
[2] "ESPAR;ESPAR0;2007 - 2008;Qualitatif;Espèce arborée;Espèce botanique d'arbre ou groupe d'espèces détaillé;"
[3] "ESPAR;ESPAR1;2009 - ;Qualitatif;Espèce arborée;Espèce botanique d'arbre ou groupe d'espèces détaillé;"
```

CET exemple montre que [7] « 3 unités se sont succédé pour la donnée `ESPAR` (unité `ESPAR` de 2005 à 2006, `ESPAR0` de 2007 à 2008 et `ESPAR1` depuis 2009). La donnée existe donc depuis 2005, mais ses unités (et modalités) ont changé. » La variable `CD_REF` est en revanche stable pour toutes les campagnes [4], on préférera donc travailler avec elle. La variable qui nous intéresse tout particulièrement est celle qui code pour l'abondance :

```
meta[substr(meta, 1, 6) == "ABOND;"]
[1] "ABOND;Indicateur d'abondance;Indicateur d'abondance-dominance d'une espèce basé sur le taux de surface couv"
[2] "ABOND;ABOND; - ;Qualitatif;Coefficient d'abondance-dominance;Taux de surface couverte par une espèce veget"
[3] "ABOND;1;0-25% présence faible;Taux de recouvrement de l'espèce inférieur à 5 % et présence faible.;"
[4] "ABOND;2;0-25% présence nette;Taux de recouvrement de l'espèce inférieur à 25 % mais présence nette.;"
[5] "ABOND;3;25-50%;Taux de recouvrement de l'espèce compris entre 25 et 50 %.;"
[6] "ABOND;4;50-75%;Taux de recouvrement de l'espèce compris entre 50% et 75 %.;"
[7] "ABOND;5;75-100%;Taux de recouvrement de l'espèce supérieur à 75%.";
```

IL S'AGIT DONC d'une variable qualitative ordonnée ayant 5 modalités. Le découpage suit les quartiles, mais pour le premier on distingue les espèces rares de celles ayant une présence nette.

LES codes taxonomiques sont documentés dans la variable `CDREF13`, on fait la même opération qu'avec le fichier `espar-cdref13.csv` pour repérer les espèces de chênes et on vérifie que l'on obtient bien la même chose.

```
cdref13 <- meta[substr(meta, 1, 8) == "CDREF13;"]
gch2 <- cdref13[grep("Chêne", cdref13)]
gqu2 <- cdref13[grep("Quercus", cdref13)]
identical(gch2, gqu2) # Sanity check
[1] TRUE
```

1.1.3 Comparaison des espèces des deux fichiers

```
gch2 <- read.table(textConnection(gch2), sep = ";")
gch2 <- gch2[, 2:4]
names(gch2) <- c("cd_ref", "lib_mix", "lib_aut")
# Sanity check
all(gch2$cd_ref %in% gch2$cd_ref)
[1] FALSE
all(gch2$cd_ref %in% gch2$cd_ref)
[1] FALSE
# Cherchons les différences
gch2[!(gch2$cd_ref %in% gch2$cd_ref), ]
  cd_ref          lib_mix          lib_aut
2 116672 Quercus coccifera (Chêne kermès) Quercus coccifera L., 1753
12 197006          Quercus (Chêne (Genre))          Quercus L., 1753
gch2[!(gch2$cd_ref %in% gch2$cd_ref), ]
  espar          lib_espar cd_ref          lib_cdref
104 08P Chêne faux-liège 116677 Quercus crenata
```

NOUS n'avons pas exactement les mêmes informations entre les deux fichiers. Le code 197006 pour *Quercus spp.* n'existe pas dans le fichier `espar-cdref13.csv`, mais cela peut sembler logique de ne pas y renseigner ce qui n'est décrit qu'au niveau du genre. En revanche, le chêne kermès (*Quercus coccifera*) n'y est pas, et le chêne faux-liège (*Quercus crenata*) n'existe pas dans `metadonnees.csv`. Pour être certain de ne perdre aucune information, on construit la table GCH qui est l'union des informations disponibles dans les deux fichiers.

```
GCH <- data.frame(list(cd_ref = unique(c(gch$cd_ref, gch2$cd_ref)))
GCH <- merge(GCH, gch, by.x = "cd_ref", by.y = "cd_ref", all.x = TRUE)
GCH <- merge(GCH, gch2, by.x = "cd_ref", by.y = "cd_ref", all.x = TRUE)
nrow(GCH)
[1] 13
```

POUR voir s'il n'y a pas de problème, on vérifie que les valeurs de `cd_ref` existent bien dans la version 13 de la nomenclature du MNHN [12].

```
taxref13 <- read.table("TAXREF_v13_2019/TAXREFv13.txt", sep = "\t",
header = TRUE, quote = "")
taxref13 <- taxref13[!duplicated(taxref13$NOM_VALIDE), ]
save(taxref13, file = "taxref13.Rda")

load("taxref13.Rda")
all(GCH$cd_ref %in% taxref13$CD_REF)
[1] TRUE

GCH <- merge(GCH, taxref13[, c("CD_REF", "NOM_VALIDE")],
by.x = "cd_ref", by.y = "CD_REF")
nrow(GCH)
[1] 13
```

ON complète « à la main » la table GCH de façon à ne plus avoir de valeurs manquantes dans les différents libellés des noms des espèces.

```
# Cas du chêne faux-liège :
GCH[GCH$cd_ref == 116677, "lib_mix"] <- "Quercus crenata (Chêne faux-liège)"
GCH[GCH$cd_ref == 116677, "lib_aut"] <- "Quercus crenata Lam., 1785"
# Cas du chêne Kermès :
GCH[GCH$cd_ref == 116672, "lib_espar"] <- "Chêne kermès"
GCH[GCH$cd_ref == 116672, "lib_cdref"] <- "Quercus coccifera"
# Cas du genre Chêne :
GCH[GCH$cd_ref == 197006, "lib_espar"] <- "Chêne (Genre)"
GCH[GCH$cd_ref == 197006, "lib_cdref"] <- "Quercus spp."
# Sanity check
any(is.na(GCH[, -2])) # Doit être FALSE
[1] FALSE
```

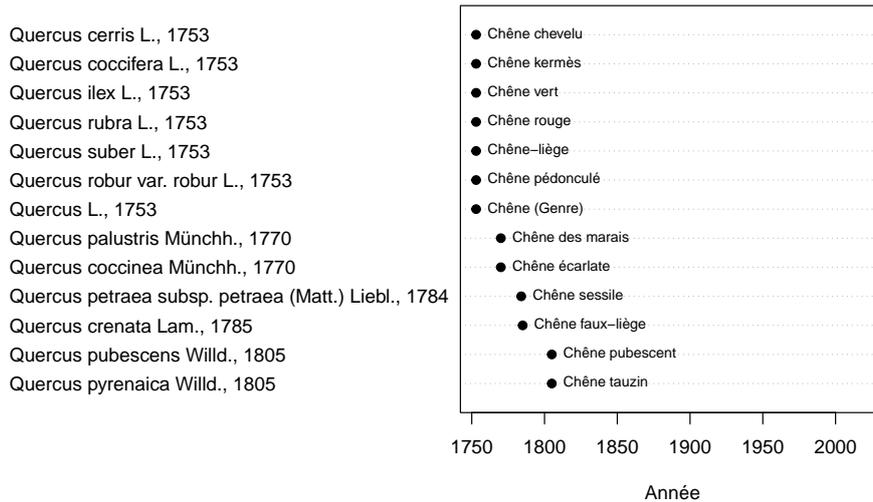
JE vérifie que les deux dernières colonnes sont identiques puis je ne conserve que les variables qui m'intéressent.

```
with(GCH, all(lib_aut == NOM_VALIDE))
[1] TRUE
GCH <- GCH[, -c(2, 7)]
```

ON peut facilement extraire la date de première description du libellé avec les auteurs. Les espèces qui nous intéressent sont décrites depuis longtemps puisque la moitié étaient déjà présentes chez LINNÉ en 1753 et que les plus récentes datent de 1805.

```
par(mar = c(5, 20, 0, 1) + 0.1, oma = c(0, 0, 2, 0))
GCH$dpp <- sapply(GCH$lib_aut, \(x) as.integer(substr(x, nchar(x) - 3, nchar(x))))
GCH <- GCH[rev(order(GCH$dpp)), ]
dotchart(GCH$dpp, pch = 19, xlim = c(min(GCH$dpp), 2023), labels = GCH$lib_aut,
         xlab = "Année")
text(GCH$dpp, 1:nrow(GCH), GCH$lib_espar, pos = 4, cex = 0.75)
title(main = "Date de première description des espèces de chênes", outer = TRUE)
```

Date de première description des espèces de chênes



1.2 Les placettes PLACETTE.csv

La géolocalisation des placettes (figure 1 page 7) est donnée dans le fichier PLACETTE.csv qui utilise un point-virgule comme séparateur des colonnes et un point comme séparateur décimal.

```
placette <- read.table("export_dataifn_2005_2021/PLACETTE.csv",
                      sep = ";", dec = ".", header = TRUE)
dim(placette)
[1] 181940 57
```

La colonne IDP est une clef d'identification *externe* des placettes, on peut vérifier que ce n'est pas une clef primaire de la table :

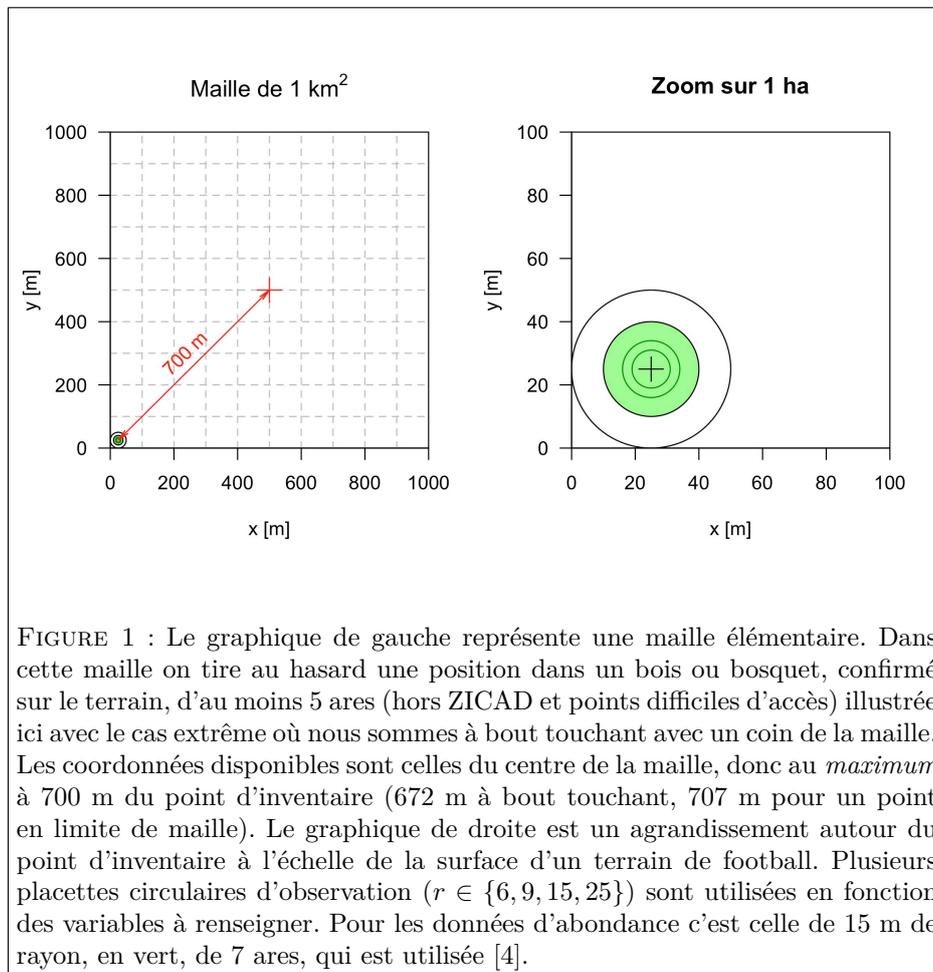
```
any(duplicated(placette$IDP))
[1] TRUE
```

En effet, une même placette peut faire l'objet d'une seconde visite, et cette information est portée dans la colonne VISITE :

```
table(placette$VISITE, useNA = "always")
 1      2 <NA>
108741 73199 0
```

On peut vérifier qu'au sein de chaque modalité de la variable VISITE on a bien unicité des identifiants des placettes :

```
with(subset(placette, VISITE == 1), any(duplicated(IDP)))
[1] FALSE
```



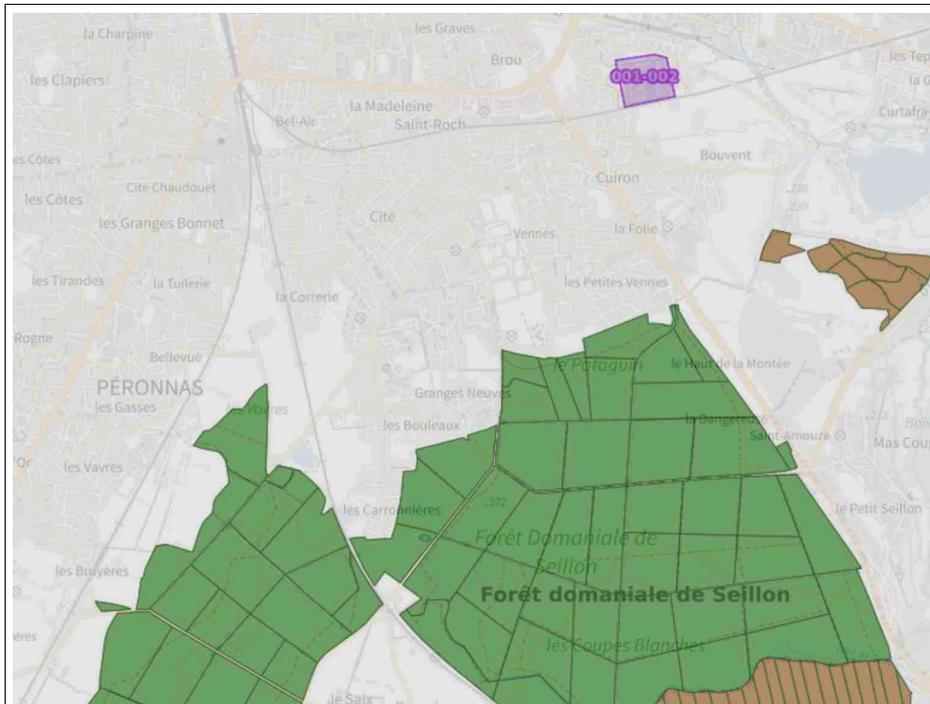


FIGURE 2 : Les ZICAD (ex ZIPVA) sont des zones interdites à la captation aérienne de données. Les coordonnées des polygones les définissant sont fixées réglementairement dans l'annexe I de l'arrêté du 2 janvier 2023 « fixant la liste des zones interdites à la captation et au traitement des données recueillies depuis un aéronef » dont on trouvera facilement la version consolidée sur le site de LEGIFRANCE si besoin est. Les principaux ministères concernés sont sans surprise ceux en charge de l'armée, de l'intérieur, de la justice et de l'énergie mais on notera deux *outliers* dans le giron du ministère de l'enseignement supérieur et de la recherche. On peut visualiser facilement les ZICAD sur le site GEOPORTAIL, par exemple ici la ZICAD 001-002 à environ 1 km au Nord de la forêt domaniale de SEILLON correspondant au centre pénitentiaire de BOURG-EN-BRESSE. Sur les ZICAD, la documentation l'inventaire forestier national dit prudemment [7] qu'« aucune visite de terrain n'y est effectuée même si certaines de ces zones relèvent sans doute de la définition de la forêt. » Cependant, si on essaye de visualiser les ZICAD à l'échelle de la France métropolitaine, on se rend compte qu'elles ne représentent qu'une fraction infinitésimale du territoire, il est donc peu vraisemblable que les ZICAD soient un facteur de nuisance pour apprécier la localisation des espèces de chênes en France métropolitaine.

```
with(subset(placette, VISITE == 2), any(duplicated(IDP)))  
[1] FALSE
```

JE pensais que toutes les placettes ayant fait l'objet d'une seconde visite devaient forcément avoir fait l'objet d'une première visite, mais j'ai trouvé une centaine d'exceptions. Comme c'est une situation très marginale (moins de 0.2 % des parcelles en seconde visite) je ne creuserai pas plus avant la question. Mais il faudra le garder en tête pour les opérations de jointure.

```
IDPV1 <- subset(placette, VISITE == 1)[, "IDP"]  
IDPV2 <- subset(placette, VISITE == 2)[, "IDP"]  
all(IDPV2 %in% IDPV1)  
[1] FALSE  
IDPV2only <- IDPV2[!(IDPV2 %in% IDPV1)]  
length(IDPV2only)  
[1] 132
```

LA nomenclature pour les codes IDP est la suivante : les 5 derniers chiffres codent pour une placette et les chiffres antérieurs pour une campagne (2005 = 0, 2006 = 1, etc.). Pour récupérer l'information, on transforme la variable IDP en une chaîne de caractères.

```
placette$IDPs <- sprintf("%07d", placette$IDP)  
placette$codecamp <- as.integer(substr(placette$IDPs, 1, 2))  
with(placette, table(CAMPAGNE, codecamp))[1:10, 1:10]  
      codecamp  
CAMPAGNE 0 1 2 3 4 5 6 7 8 9  
2005 7004 0 0 0 0 0 0 0 0 0  
2006 0 6809 0 0 0 0 0 0 0 0  
2007 0 0 7404 0 0 0 0 0 0 0  
2008 0 0 0 6739 0 0 0 0 0 0  
2009 0 0 0 0 6583 0 0 0 0 0  
2010 5824 0 0 0 0 6919 0 0 0 0  
2011 0 5747 0 0 0 0 6941 0 0 0  
2012 0 0 6429 0 0 0 0 7119 0 0  
2013 0 0 0 5780 0 0 0 0 6524 0  
2014 0 0 0 0 5665 0 0 0 0 6102
```

ON retrouve ici sur la diagonale principale les placettes ayant fait l'objet d'une première visite et avec un décalage de 5 ans celles ayant fait l'objet d'une seconde visite.

LES coordonnées géographiques données dans les colonnes XL et YL sont le centre de la maille kilométrique de la grille d'échantillonnage contenant la placette (figure 1 page 7). Les coordonnées géographiques ne devraient donc pas varier entre la première et la seconde visite, vérifions.

```
inter <- intersect(IDPV1, IDPV2)  
verif <- subset(placette, IDP %in% inter)[, c("IDP", "VISITE", "XL", "YL")]  
verif$id <- with(verif, paste(IDP, VISITE, sep = "-"))  
verif <- verif[order(verif$id), ]  
all(diff(verif$XL)[c(T, F)] == 0)  
[1] TRUE  
all(diff(verif$YL)[c(T, F)] == 0)  
[1] TRUE
```

LES coordonnées géographiques sont en LAMBERT 93, par commodité, pour pouvoir utiliser le fond de carte du paquet `maps` [10], je vais les convertir en latitude et longitude.

```

lambert93toGPS <- function(x, y){
  require(sp)
  # Définir la projection d'origine (Lambert 93)
  proj_origine <- CRS("+init=epsg:2154")
  # Définir la projection de destination (WGS 84)
  proj_destination <- CRS("+init=epsg:4326")
  # Créer un objet de type SpatialPoints avec les coordonnées en Lambert 93
  coords_origine <- SpatialPoints(matrix(c(x, y), ncol = 2),
                                     proj4string = proj_origine)
  # Transformer les coordonnées en Lambert 93 en coordonnées géographiques en latitude et longitude
  coords_destination <- spTransform(coords_origine, proj_destination)
  # Extraire les coordonnées géographiques en latitude et longitude
  lat <- coordinates(coords_destination)[, 2]
  lon <- coordinates(coords_destination)[, 1]
  return(list(lon = lon, lat = lat))
}
res <- lambert93toGPS(placette$XL, placette$YL)
placette$lon <- res$lon ; placette$lat <- res$lat

```

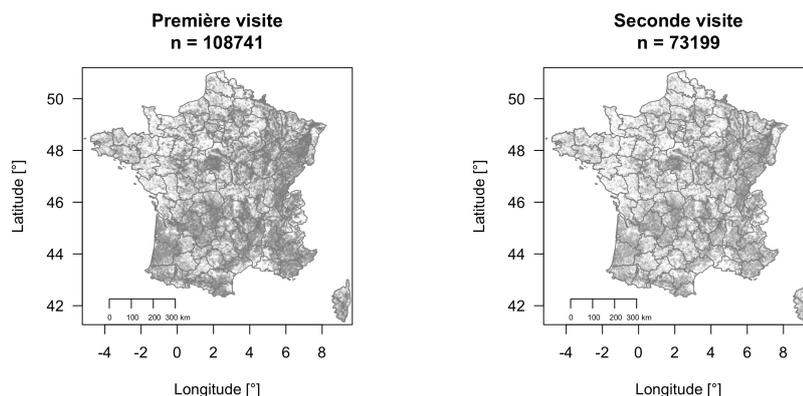
POUR se faire une première idée de la localisation des données, on pose un point gris transparent à la localisation de chaque placette.

```

mapFrance <- function(border = grey(0.5), col = "white",
  mar = c(5, 4, 4, 0), ...){
  require(maps)
  map("france", mar = mar,
      fill = TRUE, col = col, border = border, ...)
  map.axes(las = 1)
  map.scale(ratio = FALSE, cex = 0.5, relwidth = 0.2)
  title(xlab = "Longitude [°]", ylab = "Latitude [°]")
}
par(mfrow = c(1, 2), oma = c(0, 0, 2, 0))
plotplacettes <- function(visite, main, pch = 19, cex = 0.1,
  mygrey = rgb(0.5, 0.5, 0.5, 0.1)){
  mapFrance()
  nv <- sum(placette$VISITE == visite)
  title(main = paste0(main, "\nn = ", nv))
  with(subset(placette, VISITE == visite),
      points(lon, lat, pch = 19, cex = 0.1, col = mygrey))
}
plotplacettes(1, "Première visite")
plotplacettes(2, "Seconde visite")
main <- paste("Localisation des placettes\n",
  min(placette$CAMPAGNE), "-", max(placette$CAMPAGNE))
title(main = main, outer = TRUE)

```

Localisation des placettes
2005 - 2021

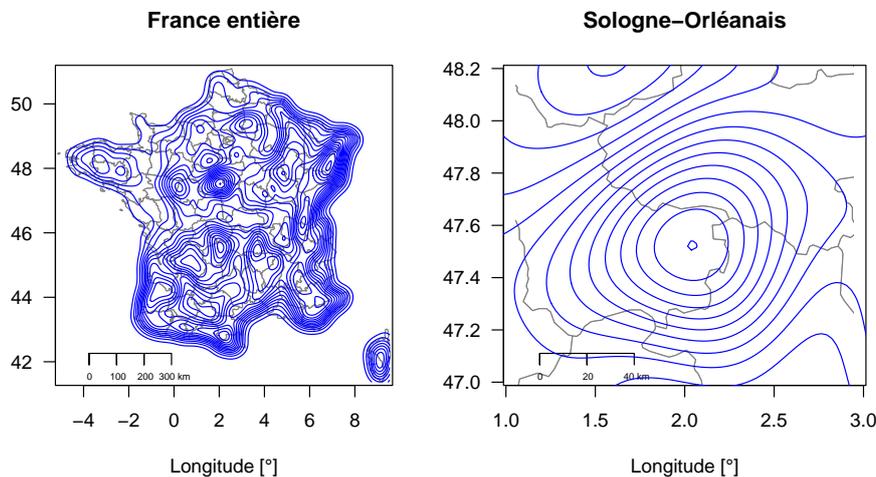


POUR explorer la distribution spatiale des placettes en première visite on peut utiliser des estimateurs de la densité locale des points grâce à la fonction `kde2d()` du paquet `MASS` [16] et utiliser des courbes de niveau pour représenter ce résumé.

```
library(MASS)
placetteV1 <- subset(placette, VISITE == 1)[ , c("lon", "lat")]
kde <- with(placetteV1, kde2d(lon, lat, n = 512))
save(placetteV1, kde, file = "kde.Rda")

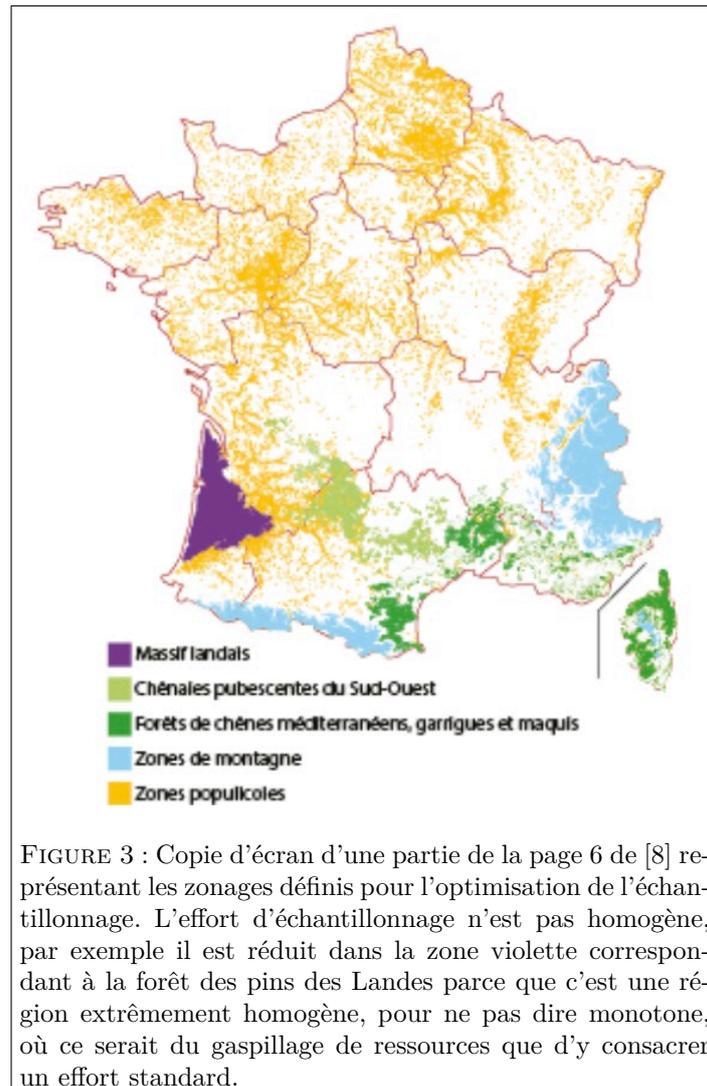
load("kde.Rda")
par(mfrow = c(1, 2), oma = c(0, 0, 2, 0), mar = c(5, 4, 2, 2))
mapFrance()
contour(kde, add = TRUE, col = "blue", drawlabels = FALSE, nlevels = 15)
title(main = "France entière")
mapFrance(xlim = c(1, 3), ylim = c(47, 48.2))
contour(kde, add = TRUE, col = "blue", drawlabels = FALSE, nlevels = 15)
title(main = "Sologne-Orléanais")
main <- paste("Densité des placettes en première visite\nn = ",
             nrow(placetteV1))
title(main = main, outer = TRUE)
```

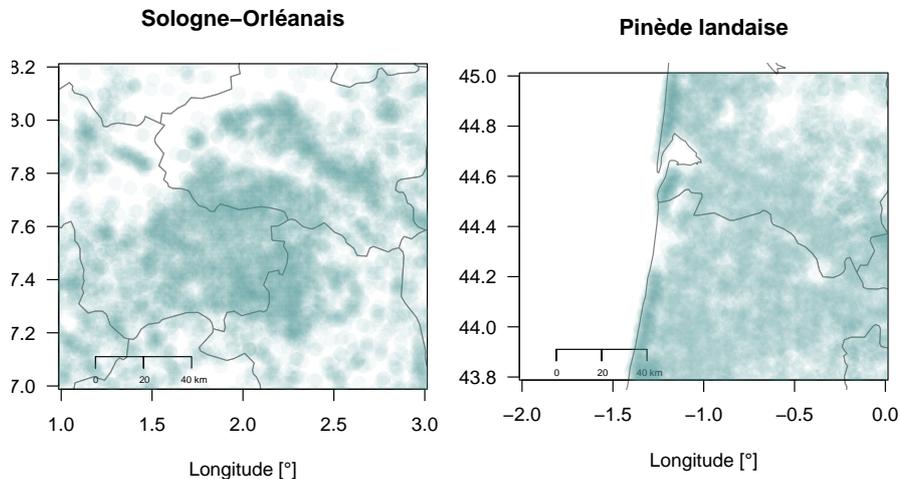
**Densité des placettes en première visite
n = 108741**



La répartition des placettes n'est à l'évidence pas homogène sur l'ensemble du territoire, globalement on observe une augmentation de la densité quand on progresse dans les terres. Quand on part d'une frontière en bord de mer, il n'y a rien à redire, en revanche la représentation est un peu trompeuse quand on part d'une frontière terrestre, on se doute bien que les arbres ne s'arrêtent pas brusquement à la frontière. La densité est faible dans les régions à forte vocation agricole, par exemple dans le grand Ouest cristallin et océanique. On reconnaît bien les massifs des Vosges et du Jura, la zone de forte densité au centre correspond aux forêts de la sylvoécocorégion B70 (Sologne-Orléanais). Attention à ne pas surinterpréter ces données car l'effort d'échantillonnage est optimisé de façon extrêmement fine (figure 3 page 12) de sorte que la surface représentée par un point d'inventaire n'est pas constante. Illustrons ce point.

```
par(mfrow = c(1, 2))
mapFrance(xlim = c(1, 3), ylim = c(47, 48.2))
mycol <- rgb(0.3, 0.6, 0.6, 0.05)
with(placetteV1, points(lon, lat, pch = 19, cex = 1.5, col = mycol))
title(main = "Sologne-Orléanais")
mapFrance(xlim = c(-2, 0), ylim = c(43.8, 45))
with(placetteV1, points(lon, lat, pch = 19, cex = 1.5, col = mycol))
title(main = "Pinède landaise")
```





ON voit ici que les massifs forestiers en Sologne-Orléanais apparaissent plus sombres que celui de la pinède landaise, non que la forêt y soit plus dense mais que l'effort d'échantillonnage y est plus conséquent. On voit aussi bien ressortir sur le graphique de droite la bande littorale atlantique où la biodiversité est plus importante : on y rencontre en particulier le chêne pédonculé (section 2.2.1 page 19), le chêne vert (section 2.3.2 page 22) et le chêne liège (section 2.4.2 page 25).

CE qui nous intéresse avant tout ici ce sont les espèces de chênes. Si on classe les zones optimisées (figure 3 page 12) dans l'ordre croissant des biais attendus pour la géolocalisation des chênes nous aurions :

- 1° le massif homogène des pins landais ;
- 2° les zones populicoles avec une forte probabilité de présence de peupleraies ;
- 3° les forêts de montagne (altitude > 1200 m et pente > 30 % ou altitude > 1700 m) ;
- 4° les forêts de chênes méditerranéens de type garrigue ou maquis, présentant un intérêt limité pour la production de bois ;
- 5° les étendues homogènes de forêts des chênaies pubescentes du Sud-Ouest.

IL faudra garder en tête que nous avons un biais de sous-représentation des chênes rencontrés en altitude, des chênes méditerranéens et surtout du chêne pubescent.

POUR faire la jointure je ne conserve dans la table `geoloc` que la clef d'identification et les coordonnées géographique des placettes. Je conserve toutes les données de la première visite et celles qui n'ont eu qu'une seconde visite. Je vérifie que j'ai conservé autant de données qu'il y a d'identifiants uniques dans le fichier de départ.

```
geoloc <- subset(placette, VISITE == 1 | IDP %in% IDPV2only)[ , c("IDP", "lon", "lat")]
any(duplicated(geoloc$IDP))
```

```
[1] FALSE
nrow(geoloc) == length(unique(placette$IDP)) # Sanity check
[1] TRUE
```

NOUS avons des informations pour 108873 placettes représentatives (hors optimisation) chacune d'une maille de 1 km². La surface forestière en France métropolitaine de nos jours est de l'ordre de 17.1 10⁶ ha, soit 17.1 10⁴ km². Les mailles échantillonnées représentent donc environ 64 % de la surface forestière totale. C'est un minorant puisque qu'une placette peut être représentative de plus de 1 km² en cas d'optimisation. On a donc au final une *très très* belle couverture du territoire forestier métropolitain.

1.3 La flore FLORE.CSV

LE code des espèces est donné dans la colonne CD_REF, nous avons déjà récupéré ceux qui nous intéressent à partir du fichier `metadonnees.csv` (section 1.1.2 page 4) dans la table GCH. Il nous suffit donc de sélectionner les placettes où des chênes ont été observés. Je ne conserve que les quatre premières colonnes de la table `flore` parce que la dernière n'est pas documentée.

```
flore <- read.table("export_dataifn_2005_2021/FLORE.csv",
                  sep = ";", dec = ".", header = TRUE)
save(flore, file = "flore.Rda")

load("flore.Rda")
chênes <- subset(flore[, 1:4], CD_REF %in% GCH$cd_ref)
nrow(chênes) # 43290
[1] 43290
range(chênes$CAMPAGNE)
[1] 2005 2021
head(chênes)
  CAMPAGNE IDP CD_REF ABOND
44      2005  19 139584     2
66      2005  23 139596     4
162     2005  47 139584     2
163     2005  47 139596     2
173     2005  54 139584     5
212     2005  69 139584     5
```

NOUS avons donc 43290 placettes où des chênes ont été observés entre 2005 et 2021. Comme le code CD_REF n'est pas très parlant, on fait une première opération de jointure⁹ avec la table GCH pour récupérer le nom scientifique et vernaculaire des espèces.

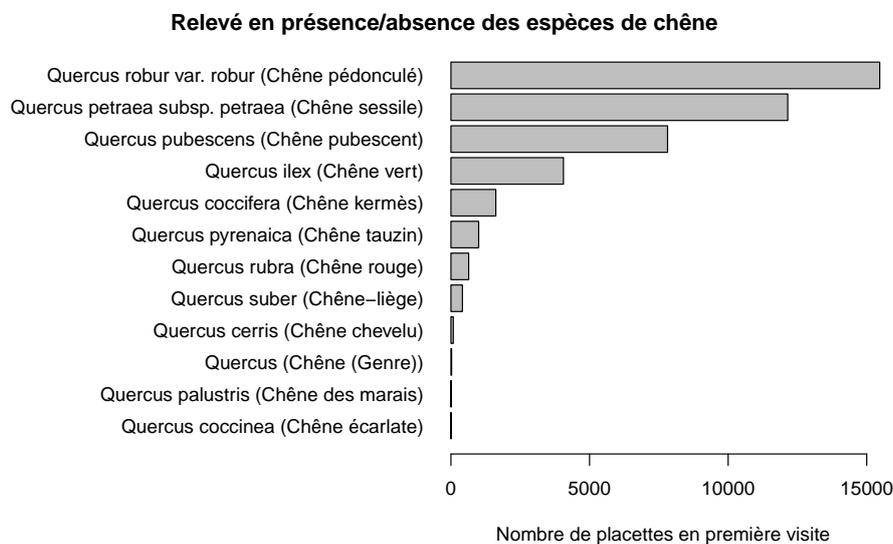
```
chênes <- merge(chênes, GCH,
               by.x = "CD_REF", by.y = "cd_ref")
nrow(chênes) # Sanity check
[1] 43290
head(chênes)
  CD_REF CAMPAGNE   IDP ABOND   lib_espar   lib_cdref
1 116670   2018 1330234     2 Chêne chevelu Quercus cerris
2 116670   2018 1311537     2 Chêne chevelu Quercus cerris
3 116670   2006 149844     2 Chêne chevelu Quercus cerris
4 116670   2005 12434     1 Chêne chevelu Quercus cerris
5 116670   2005 12928     4 Chêne chevelu Quercus cerris
6 116670   2016 1113484     1 Chêne chevelu Quercus cerris
      lib_mix      lib_aut dpp
```

⁹Le logiciel  n'est pas un SGBD mais on peut facilement faire toutes les opérations de l'algèbre relationnelle, voir <https://pbil.univ-lyon1.fr/R/pdf/tdr19.pdf>

```
1 Quercus cerris (Chêne chevelu) Quercus cerris L., 1753 1753
2 Quercus cerris (Chêne chevelu) Quercus cerris L., 1753 1753
3 Quercus cerris (Chêne chevelu) Quercus cerris L., 1753 1753
4 Quercus cerris (Chêne chevelu) Quercus cerris L., 1753 1753
5 Quercus cerris (Chêne chevelu) Quercus cerris L., 1753 1753
6 Quercus cerris (Chêne chevelu) Quercus cerris L., 1753 1753
```

NOUS pouvons maintenant compter le nombre de placettes où chaque espèce de chêne a été observée, quelle que soit son abondance.

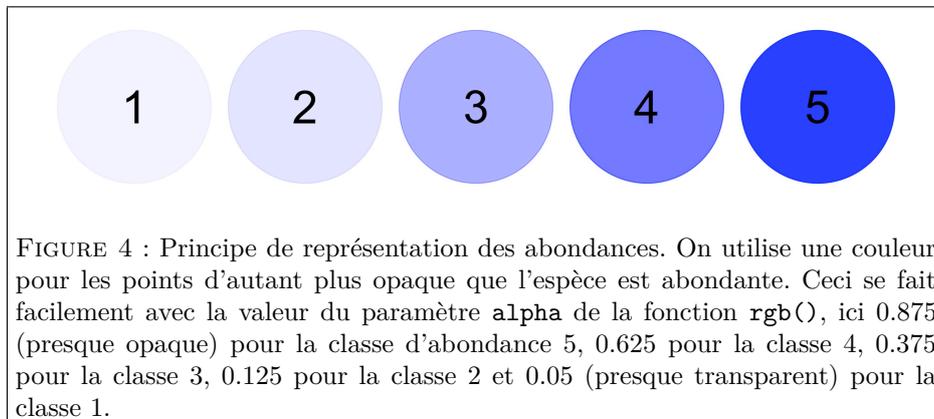
```
tabch <- with(chênes, sort(table(lib_mix)))
par(mar = c(5, 20, 0, 1) + 0.1, oma = c(0, 0, 2, 0))
barplot(tabch, horiz = TRUE, las = 1,
        xlab = "Nombre de placettes en première visite")
title(main = "Relevé en présence/absence des espèces de chêne", outer = TRUE)
```



LE chêne sessile et le chêne pédonculé dominent largement le haut du tableau, leur cas méritera que l'on s'y attarde. Le chêne pubescent et le chêne vert sont deux outsiders, n'oublions pas que le chêne pubescent est sous-échantillonné. Le chêne chevelu, le chêne des marais et le chêne écarlate semblent assez peu représentés. On note qu'il n'y a que 24 placettes qui ne sont documentées qu'au niveau du genre, on ne les considérera plus dans la suite. Mais cette vision en simple présence/absence lors de l'observation des placette est très dégradée puisque l'on ne tient pas compte de l'abondance. Construisons la table de contingence croisant les espèces et leur abondances.

```
chênes <- subset(chênes, CD_REF != 197006) # Censure du genre
tabab <- with(chênes, table(lib_espar, ABOND))
tabab[rev(order(tabab[,5])), ]
```

lib_espar	ABOND				
	1	2	3	4	5
Chêne sessile	4985	4473	1290	824	578
Chêne pubescent	2881	3363	733	458	380
Chêne pédonculé	6894	6292	1249	682	352
Chêne vert	1016	1858	727	310	148
Chêne kermès	379	538	361	202	139
Chêne rouge	338	217	50	18	17
Chêne chevelu	33	30	7	9	6
Chêne tauzin	551	398	40	11	1
Chêne-liège	174	164	53	23	0
Chêne écarlate	2	1	0	0	0
Chêne des marais	5	4	0	2	0



ON voit que le chêne sessile est celui qui est le plus fréquent en situation de forte abondance (modalités 4 et 5) alors que le chêne pédonculé est le champion en situation de faible abondance (modalités 1 et 2).

2 Localisation des chênes

POUR pouvoir localiser les observations des chênes sur une carte, je fais la jointure entre la table `chênes` qui me donne les relevés d'abondance par placette et la table `geoloc` qui me donne la géolocalisation d'icelles.

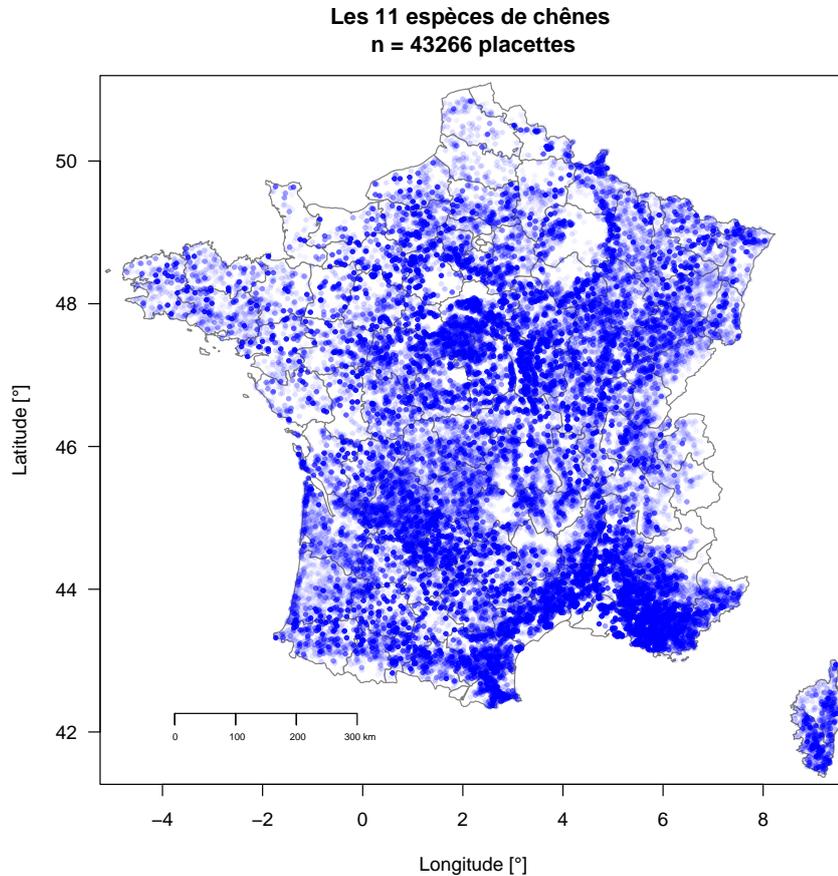
```
chênes <- merge(chênes, geoloc, by.x = "IDP", by.y = "IDP")
```

2.1 Principe des cartes

JE vais utiliser deux types de représentations. Quel que soit le type utilisé il ne faut pas oublier que l'optimisation de l'échantillonnage fait qu'il peut y avoir des biais dans ces représentations.

DANS une première représentation, on pose un point d'autant plus opaque que l'espèce est abondante (figure 4 page 16). Voici le résultat toutes espèces confondues.

```
mapchene <- fonction(dta, cdref, nom = "", col = c(0, 0, 1), cex = 0.5,
  alpha = c(0.05, 0.125, 0.375, 0.625, 0.875), ...){
  mapFrance(...)
  tmp <- subset(dta, CD_REF %in% cdref)
  if(nom == "") nom <- tmp[1, "lib_mix"]
  cols <- rgb(col[1], col[2], col[3], alpha)
  points(tmp$lon, tmp$lat, pch = 19, cex = cex, col = cols[tmp$ABOND])
  title(main = paste(nom, "\nn =", nrow(tmp), "placettes"))
}
allspecies <- unique(chênes$CD_REF)
mapchene(chênes, allspecies, "Les 11 espèces de chênes")
```



COMME on peut le constater, on rencontre des chênes sur la quasi-totalité du territoire métropolitain. Le maillage est extrêmement serré, ainsi, même si l'effort d'échantillonnage n'est pas homogène à cause de son optimisation, du point de vue de la géolocalisation il est peu probable qu'une espèce puisse passer entre les mailles du filet.

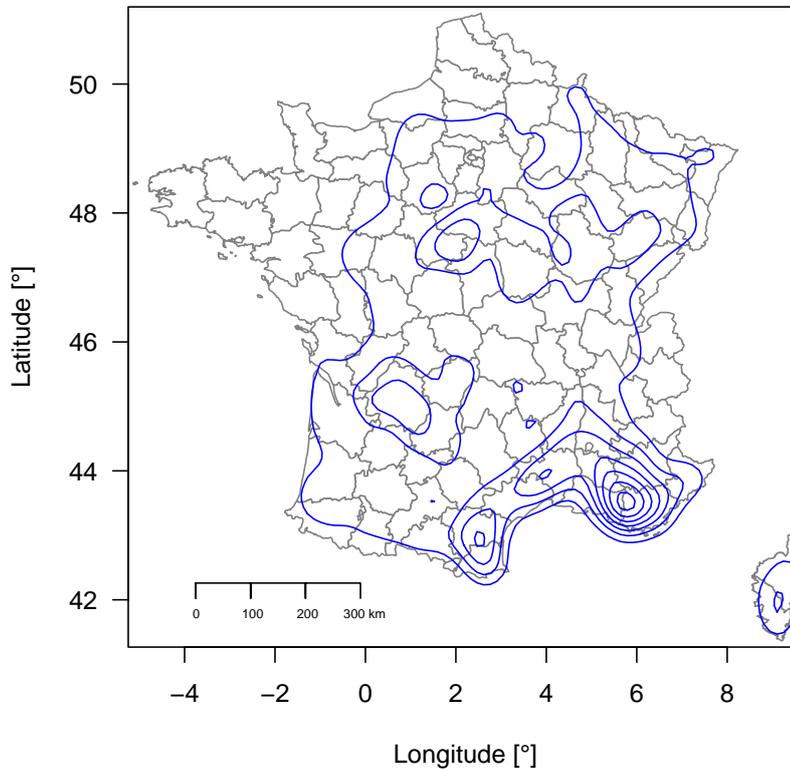
LE deuxième type de représentation est du type courbe de niveau. Pour tenir compte des abondances on répète chaque point autant de fois que sa modalité d'abondance.

```

curvechene <- function(dta, cdref, nom = "", col = rgb(0, 0, 1), n = 100, ...){
  require(MASS)
  mapFrance(...)
  tmp <- subset(dta, CD_REF %in% cdref)
  if(nom == "") nom <- tmp[1, "lib_mix"]
  title(main = paste(nom, "\nn =", nrow(tmp), "placettes"))
  x <- rep(tmp$lon, times = tmp$ABOND)
  y <- rep(tmp$lat, times = tmp$ABOND)
  kde <- kde2d(x, y, n = n)
  contour(kde, add = TRUE, col = col, drawlabels = FALSE)
}
curvechene(chênes, allspecies, "Les 12 espèces de chênes")

```

Les 12 espèces de chênes n = 43266 placettes



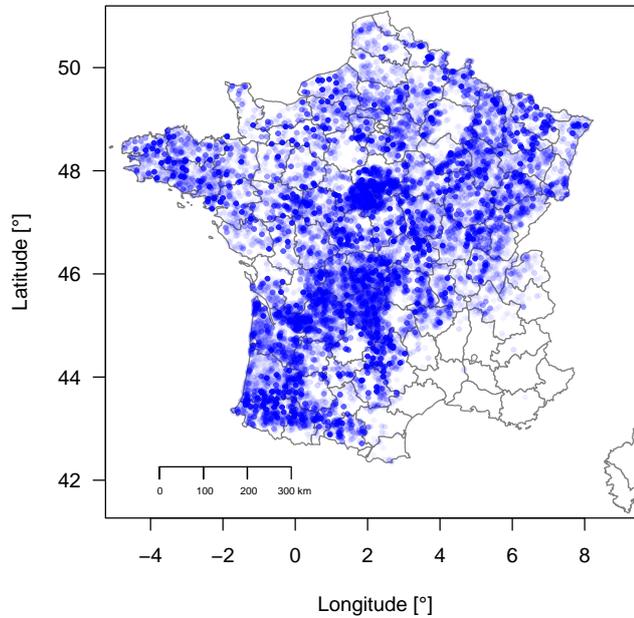
CE qui ressort ici, comme on le verra en détail par la suite, c'est la plus grande biodiversité des chênes sur le pourtour méditerranéen puisque l'on y trouve du chêne pubescent (section 2.3.1 page 21), du chêne vert (section 2.3.2 page 22), du chêne kermès (section 2.4.1 page 24), du chêne-liège (section 2.4.2 page 25), et du chêne écarlate (section 2.5.3 page 27).

2.2 Les rois de la forêt

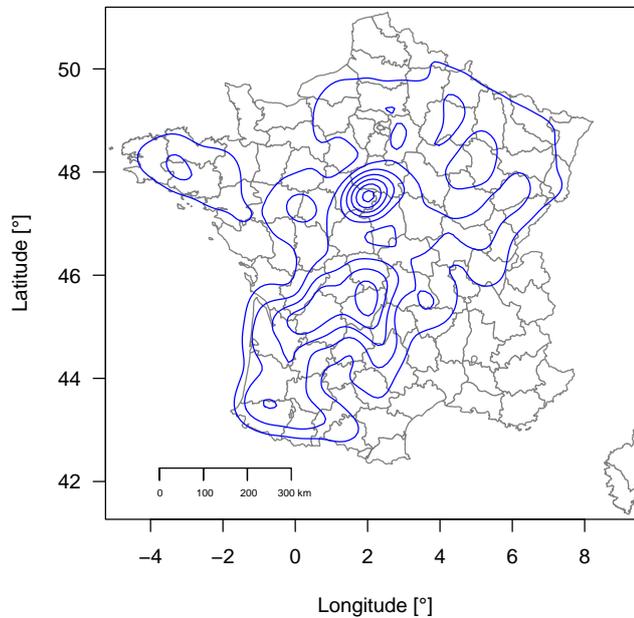
LE chêne sessile et le chêne pédonculé entretiennent une relation tout à fait singulière dont on pourra lire une description saisissante en français [13] ou en anglais [14]. Ce sont deux espèces parfaitement autonomes mais capables de produire des hybrides de façon asymétrique : le pollen de chêne sessile (aussi appelé chêne mâle) peut féconder les ovules du chêne pédonculé (aussi appelé chêne femelle) mais la réciproque n'est que rarement vraie. Par croisements successifs, au bout de quelques générations d'hybrides successives, le chêne sessile finit par être « régénéré » tout en acquérant le génome cytoplasmique, à hérédité maternelle, du chêne pédonculé. Le chêne pédonculé est une espèce pionnière capable de coloniser des milieux ouverts alors que le chêne sessile consolide les bases arrières par son comportement social.

2.2.1 Chêne pédonculé

Quercus robur var. robur (Chêne pédonculé)
n = 15469 placettes

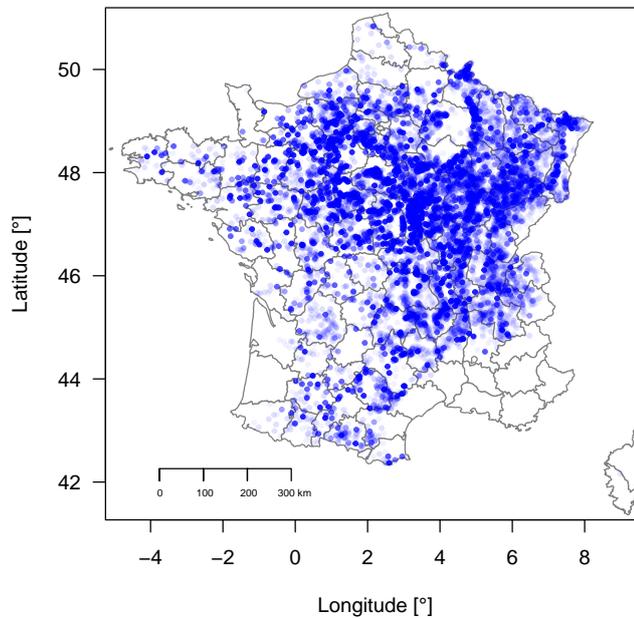


Quercus robur var. robur (Chêne pédonculé)
n = 15469 placettes

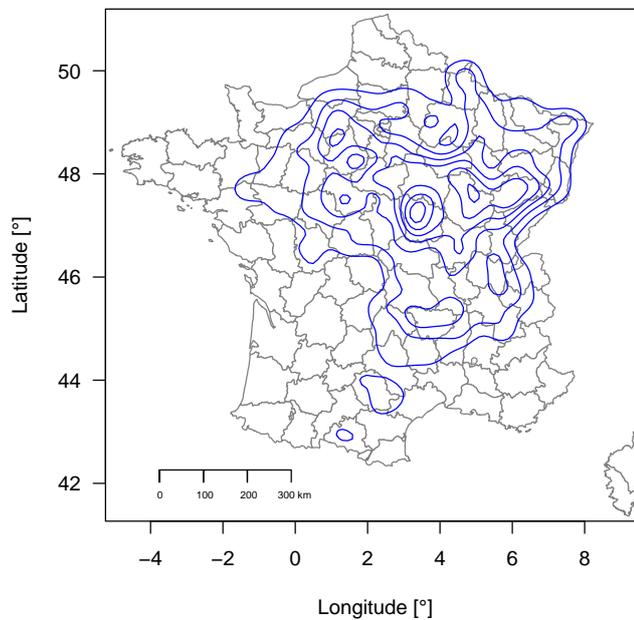


2.2.2 Chêne sessile

Quercus petraea subsp. petraea (Chêne sessile)
n = 12150 placettes

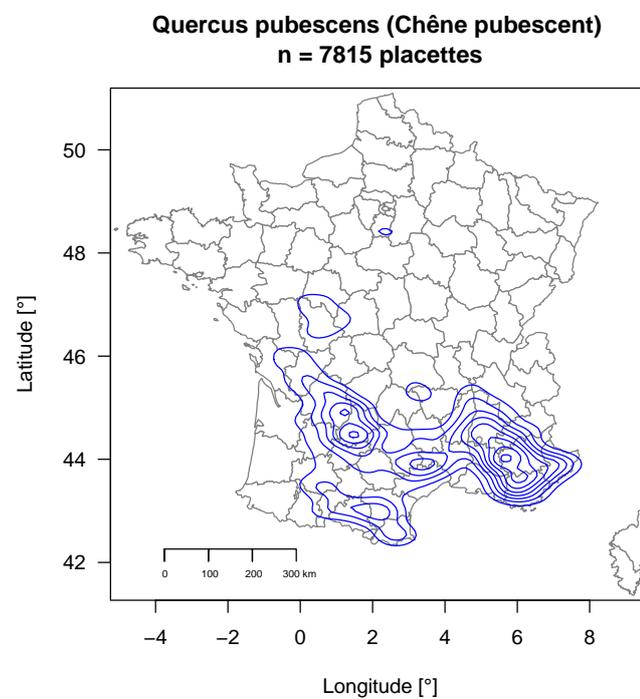
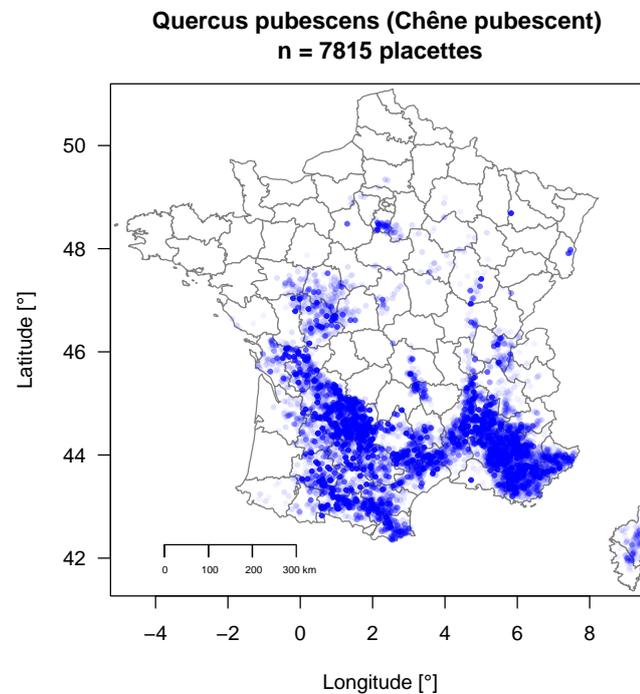


Quercus petraea subsp. petraea (Chêne sessile)
n = 12150 placettes



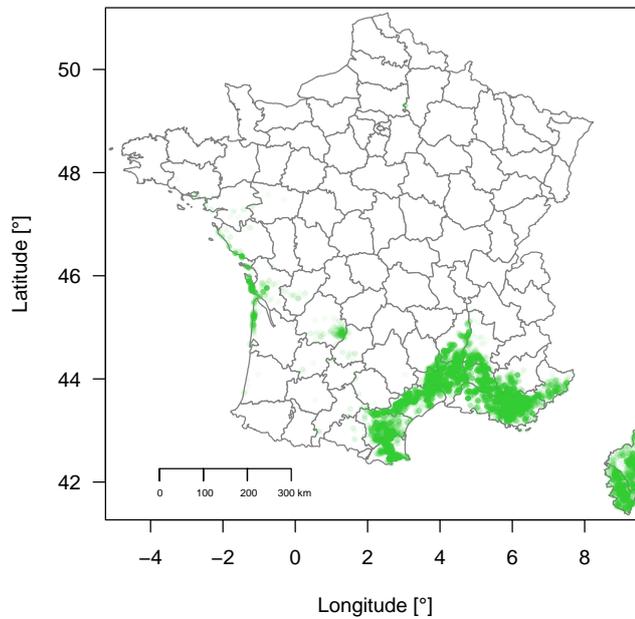
2.3 Les outsiders

2.3.1 Chêne pubescent

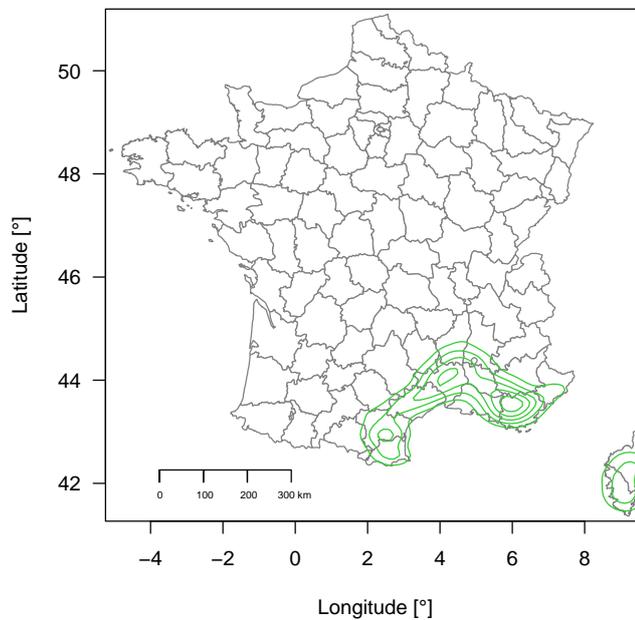


2.3.2 Chêne vert

Quercus ilex (Chêne vert)
n = 4059 placettes

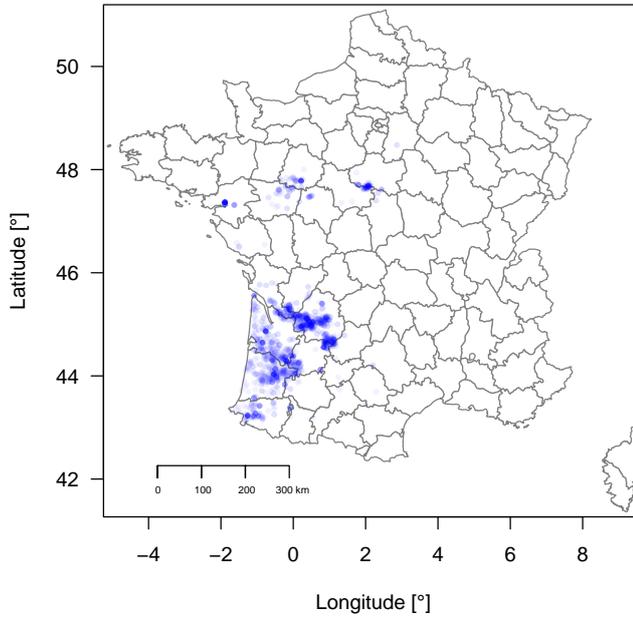


Quercus ilex (Chêne vert)
n = 4059 placettes

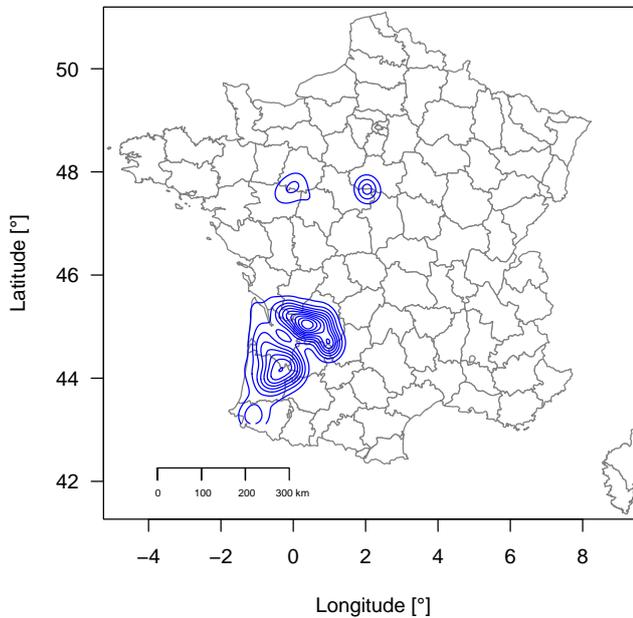


2.3.3 Chêne tauzin

Quercus pyrenaica (Chêne tauzin)
n = 1001 placettes

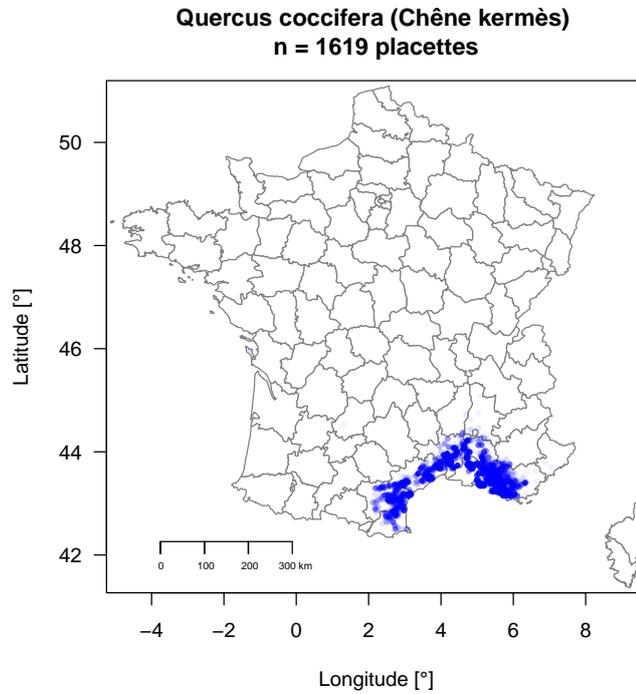


Quercus pyrenaica (Chêne tauzin)
n = 1001 placettes

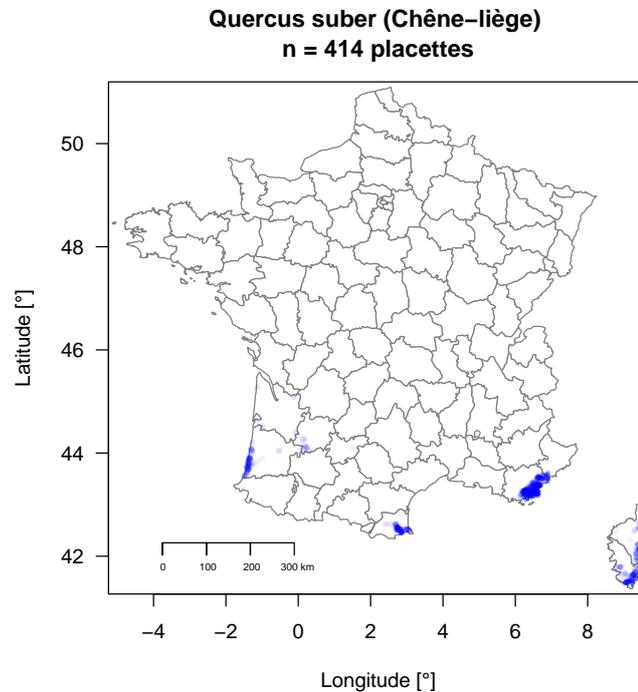


2.4 Les méridionaux

2.4.1 Chêne kermès



2.4.2 Chêne-liège



2.4.3 Chêne faux-liège

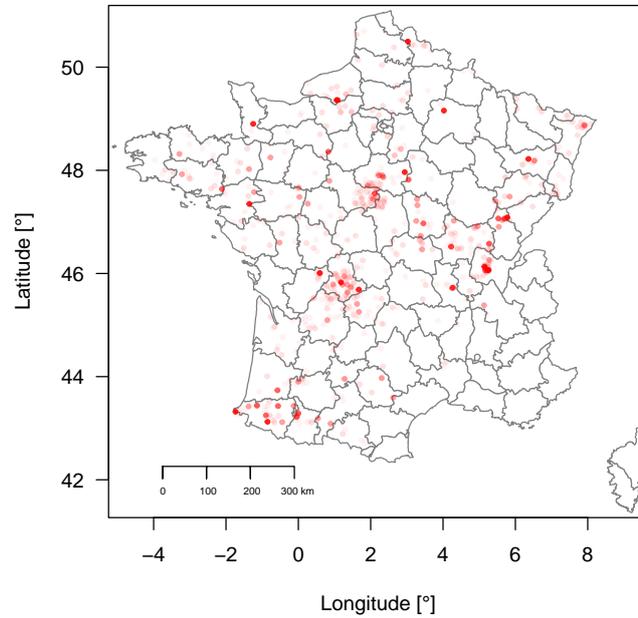
LE chêne faux-liège (*Quercus crenata*) n'est documenté sur aucune placette. Cette espèce était présente dans le fichier `espar-cdref13.csv` mais pas dans le fichier `metadonnees.csv`.

2.5 Espèces introduites récemment

2.5.1 Chêne rouge

C'EST une espèce importée d'Amérique que l'on trouve de façon sporadique sur tout le territoire du chêne sessile et pédonculé. Ce n'est pas une espèce dont on cherche le maintien, elle est considérée comme invasive [17, 15] en Europe, par exemple le plan d'aménagement de la forêt domaniale de SEILLON [11] qui contient 3 % de chênes de rouges prévoit de les éliminer à long terme.

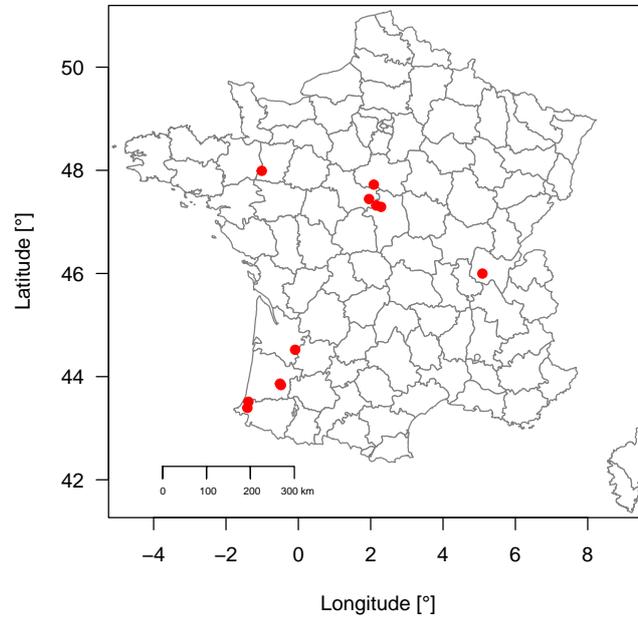
Quercus rubra (Chêne rouge)
n = 640 placettes



2.5.2 Chêne des marais

CETTE espèce, qui n'est pas spécialement hygrophile malgré ce que son nom pourrait suggérer, à été importée d'Amérique. Elle n'est documentée que sur 11 placettes.

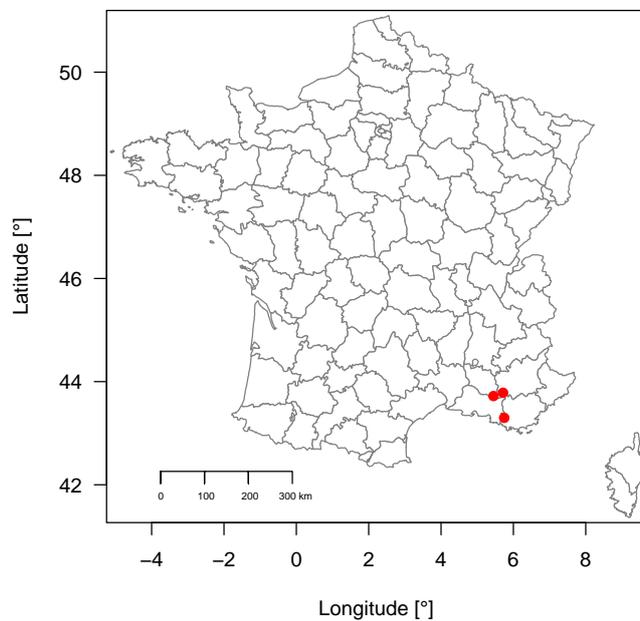
Quercus palustris (Chêne des marais)
n = 11 placettes



2.5.3 Chêne écarlate

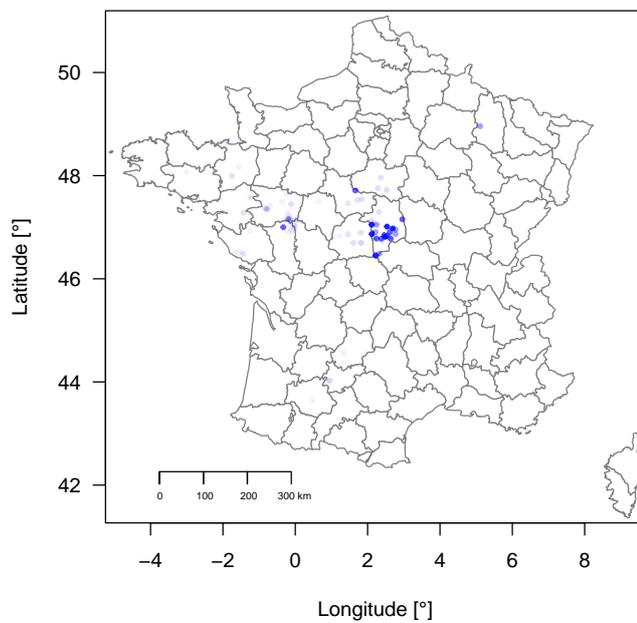
CETTE espèce importée d'Amérique n'est documentée que sur 3 placettes en région méditerranéenne. Elle est souvent confondue avec le chêne rouge et le chêne des marais.

Quercus coccinea (Chêne écarlate)
n = 3 placettes



2.5.4 Chêne chevelu

Quercus cerris (Chêne chevelu)
n = 85 placettes



Références

- [1] Anonymous. Documentation des données brutes de l'inventaire forestier mises en ligne sur DATAIFN. fichier : `BOIS_MORT.csv`. Technical report, Service de l'information statistique forestière et environnementale and Institut National de l'Information Géographique et Forestière, 2022. 3 pp. Version 1.0.
- [2] Anonymous. Documentation des données brutes de l'inventaire forestier mises en ligne sur DATAIFN. fichier : `COUVERT.csv`. Technical report, Service de l'information statistique forestière et environnementale and Institut National de l'Information Géographique et Forestière, 2022. 3 pp. Version 2.0.
- [3] Anonymous. Documentation des données brutes de l'inventaire forestier mises en ligne sur DATAIFN. fichier : `ECOLOGIE.csv`. Technical report, Service de l'information statistique forestière et environnementale and Institut National de l'Information Géographique et Forestière, 2022. 19 pp. Version 2.0.
- [4] Anonymous. Documentation des données brutes de l'inventaire forestier mises en ligne sur DATAIFN. fichier : `FLORE.csv`. Technical report, Service de l'information statistique forestière et environnementale and Institut National de l'Information Géographique et Forestière, 2022. 3 pp. Version 2.0.
- [5] Anonymous. Documentation des données brutes de l'inventaire forestier mises en ligne sur DATAIFN. fichier : `HABITAT.csv`. Technical report, Service de l'information statistique forestière et environnementale and Institut National de l'Information Géographique et Forestière, 2022. 4 pp. Version 2.0.
- [6] Anonymous. Documentation des données brutes de l'inventaire forestier mises en ligne sur DATAIFN. fichier : `PLACETTE.csv`. Technical report, Service de l'information statistique forestière et environnementale and Institut National de l'Information Géographique et Forestière, 2022. 12 pp. Version 2.0.
- [7] Anonymous. Données brutes de l'inventaire forestier mises en ligne sur DATAIFN. Technical report, Service de l'information statistique forestière et environnementale and Institut National de l'Information Géographique et Forestière, 2022. 8 pp. Version 2.0.
- [8] Anonymous. Méthodologie pour bien comprendre les résultats publiés 2017-2021. Technical report, Institut National de l'Information Géographique et Forestière, 2022. 39 pp.
- [9] Anonymous. Documentation des données brutes de l'inventaire forestier mises en ligne sur DATAIFN. fichier : `ARBRE.csv`. Technical report, Service de l'information statistique forestière et environnementale and Institut National de l'Information Géographique et Forestière, 2023. 13 pp. Version 2.1.

- [10] Original S code by Richard A. Becker, Allan R. Wilks. R version by Ray Brownrigg. Enhancements by Thomas P Minka, and Alex Deckmyn. *maps : Draw Geographical Maps*, 2022. R package version 3.4.1.
- [11] S. Dumas and D. Pivot. Forêt domaniale de SEILLON. surface : 613,68 ha. révision d'aménagement forestier 2009-2028. Technical report, Office national des Forêts, Direction territoriale Rhône-Alpes, Agence Ain Loire Rhône, 2013. 65 pp.
- [12] O. Gargominy, S. Terceirie, C. Régnier, T. Ramage, P. Dupont, P. Daszkiewicz, and L. Poncet. TAXREF, référentiel taxonomique pour la France : méthodologie, mise en œuvre et diffusion. Technical report, Rapport PatriNat (OFB-CNRS-MNHN), Muséum national d'Histoire naturelle, Paris, 2022. 47 pp.
- [13] A. Kremer, R. Petit, and A. Ducouso. Biologie évolutive et diversité génétique des chênes sessile et pédonculé. *Revue forestière française*, 54 :111–130, 2002.
- [14] R.J. Petit, C. Bodénès, A. Ducouso, G. Roussel, and A. Kremer. Hybridization as a mechanism of invasion in oaks. *New Phytologist*, 161(1) :151–164, 2004.
- [15] M. Stanek, P. Kushwaha, K. Murawska-Wlodarczyk, A.M. Stefanowicz, and A. Babst-Kostecka. *Quercus rubra* invasion of temperate deciduous forest stands alters the structure and functions of the soil microbiome. *Geoderma*, 430 :116328, 2023.
- [16] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Springer, New York, fourth edition, 2002.
- [17] B. Woziwoda, D. Kopec, and J. Witkowski. The negative impact of intentionally introduced *Quercus rubra* l. on a forest community. *Acta societatis botanicorum Poloniae*, 83(1), 2014.