

# Model Selection Report

Loïc Pages

2025-03-28

## Introduction

### IPM

$$N(y, t+1) = \int N(x, t)(F_a(x, y, t) + P_a(x, y, t))dx$$

avec  $x$  la taille à  $t$ ,  $y$  la taille à  $t+1$ ,  $a$  l'âge et  $N$  la taille de la population

La probabilité qu'un individu d'âge  $a$  et de taille  $x$  au temps  $t$  devienne un individu d'âge  $a+1$  et de taille  $y$  à  $t+1$  est :

$$P_a(x, y, t) = s_a(x)(1 - f_a(x))G_a(x, y)$$

et la densité d'individus de taille  $y$  à  $t$  et d'âge 1 (plantules) issus d'un individu de taille  $x$  et d'âge  $a$  à  $t$  est :

$$F_a(x, y, t) = f_a(x)C_a(x)w(y)Estb$$

### Sélection de modèles

Méthode utilisée: Créer l'ensemble des combinaisons possibles d'effets fixes et aléatoires. Puis utiliser le package spaMM pour fit les modèles. On calcule ensuite leur AIC (ou leur BIC) qu'on compare entre eux. On obtient un classement des (5) meilleurs modèles pour chaque trait d'histoire de vie, pour l'AIC et pour le BIC.

Paramètres à modéliser :

Survival Probability

Flowering Probability

Growth

Fecundity (Number of capitula)

Seedling Size Distribution

### Initialisation

```
rm(list=ls())  
library(knitr)  
library(spaMM)  
library(tidyverse)  
library(splines)
```

```

library(foreach)
library(doParallel)

setwd("/media/loic/Commun/0Travail/Stage 2025 ISEM/Models")

centauree_data <- read.csv("donneesIPM_short.csv")
centauree_data_complet <- read.csv("donneesIPM.csv")

#Supprimer plantes dont l'age est inconnu
centauree_data <- centauree_data[!is.na(centauree_data$age0), ]

#Ajouter valeur de l'age à t+1
centauree_data$age1 <- ifelse(centauree_data$Stage1=="V",centauree_data$age0+1,NA)

#Forcer l'age maximal à 8
centauree_data$age0[centauree_data$age0 > 8] <- 8
centauree_data$age1[centauree_data$age1 > 8] <- 8

spaMM.options(separation_max=70)

```

## Mise en place

### Survival Probability

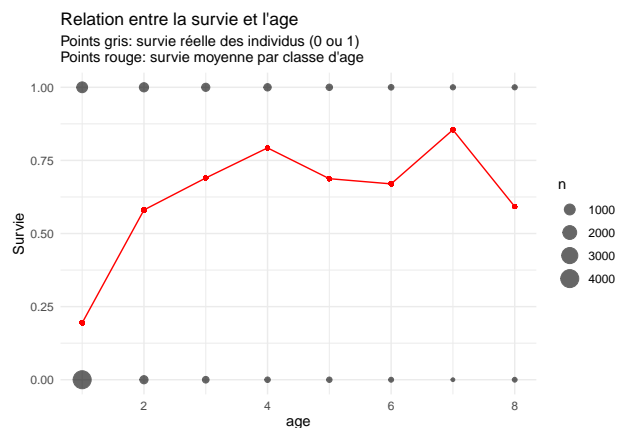
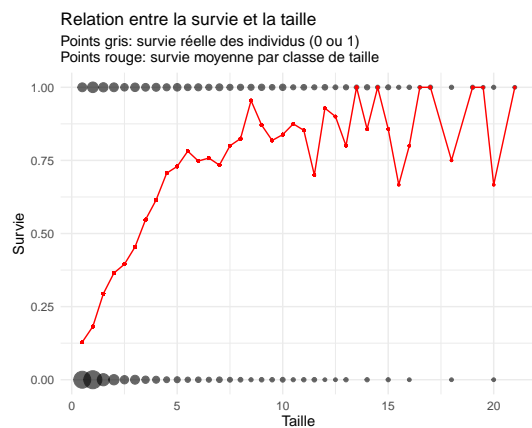
Pour le probabilité de survie, on a divisé le jeu de données en deux:

- 1- les plantules (individus d'âge 1)
- 2- les rosettes (individus d'âge 2 ou plus)

(La somme des AIC des meilleurs modèles de survie des plantules et des rosettes est inférieure à l'AIC des modèles de survie comprenant tous les âges)

On utilise un modèle binomial pour fit les données de survie (0 ou 1).

Affichage des données brutes de la survie pour tous les âges.

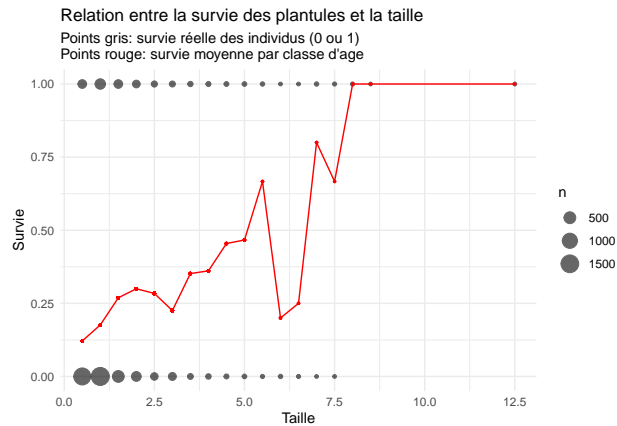


### Age 1 - Plantules

Pour la survie des plantules, les effets explorés sont :

fixe : taille (polynomes et splines)

aléatoire : population, année (effet sur l'intercept ou la taille)

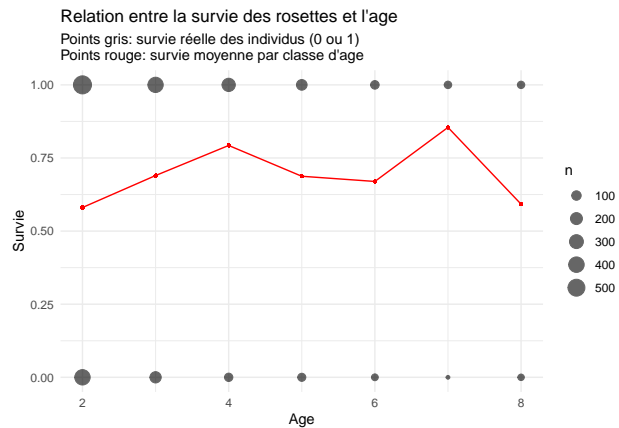
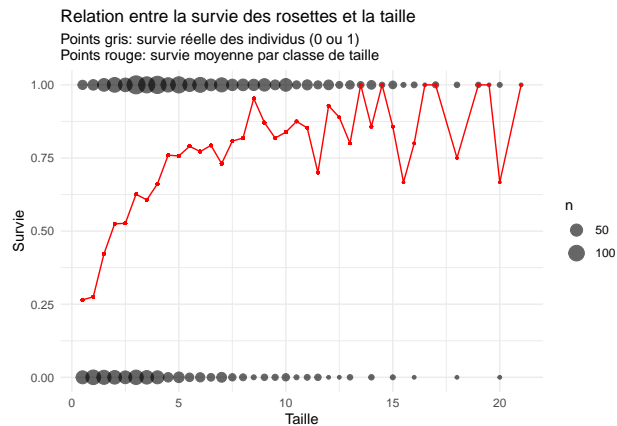


## Age +2 - Rosettes

Pour la survie des rosettes, les effets explorés sont :

fixe : taille (polynomes et splines) et age (polynome et splines)

aléatoire : population, année (effet sur l'intercept ou la taille), hétérogénéité individuelle (intercept)



## Flowering Probability

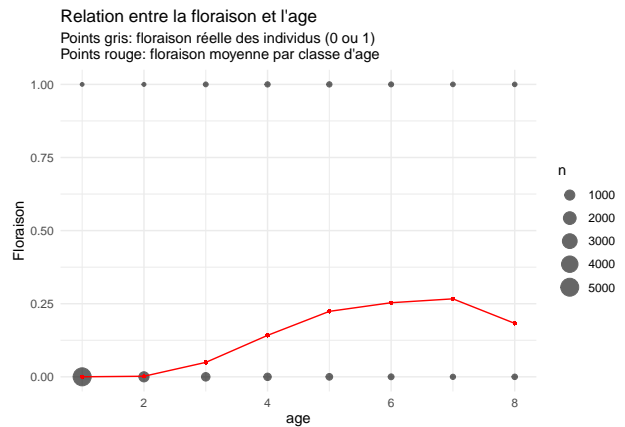
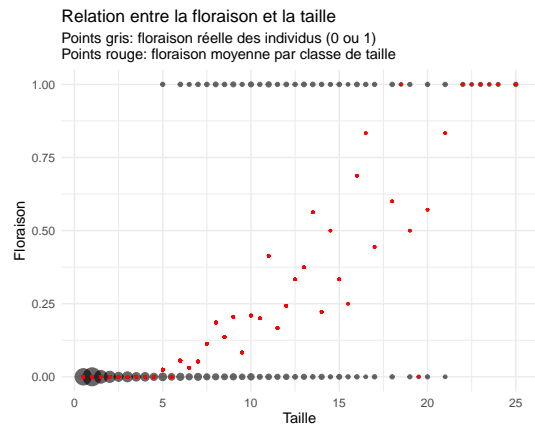
On utilise un modèle binomial pour fit les données de floraison (0 ou 1).

Pour les modèles de floraison, les possibilités de combinaisons d'effets fixes et aléatoires ont été réduit pour simplifier les calculs de sélection.

Les effets d'hétérogénéité entre individus ont été retirés par soucis de temps de calcul et de convergence des modèles.

Les effets fixes de l'age ont été simplifiés en retirant les splines.

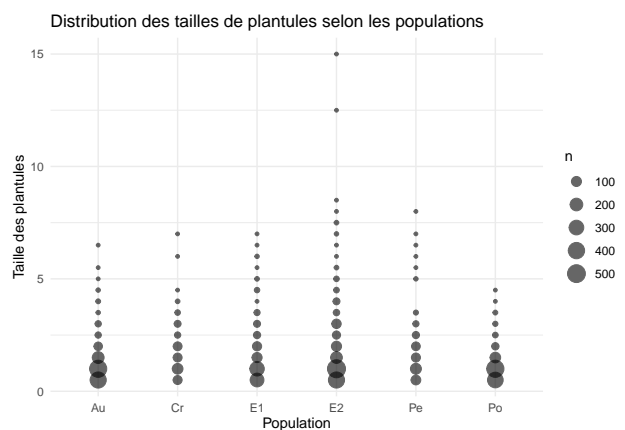
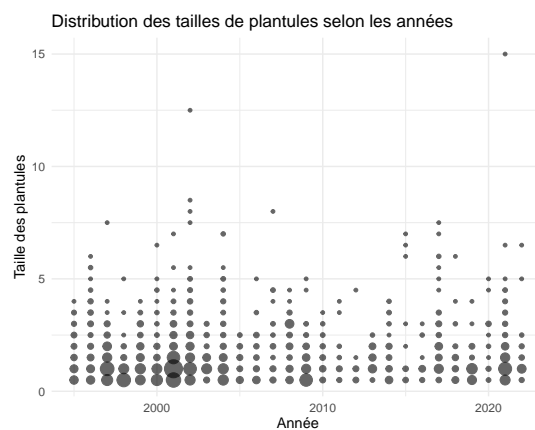
D'autres combinaisons (notamment contenant les splines de la taille) ont été retirés par un algorithme détectant si le modèle subit des effets de séparation qui empêchent la convergence du modèle.



## Seedl Size

Pour fit les données de taille de plantules, on utilise un modèle qui suit une distribution log-Gamma.

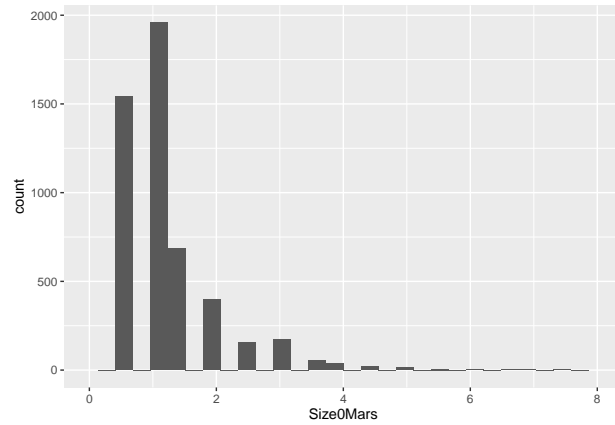
On y fait varier les effets aléatoires années, population et l'interaction année:population.



```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```

```
## Warning: Removed 3 rows containing non-finite outside the scale range
## ('stat_bin()').
```

```
## Warning: Removed 2 rows containing missing values or values outside the scale range
## ('geom_bar()').
```

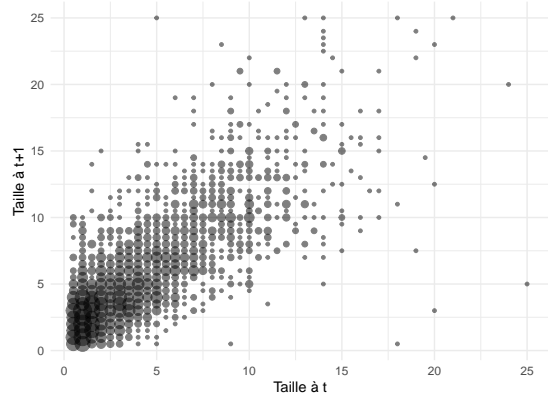


## Growth

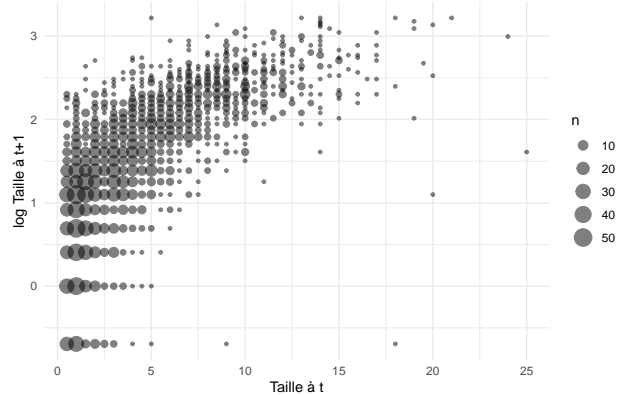
Pour la croissance des plantes, on a choisi de fit le log de la taille à l'année  $t + 1$  en fonction des effets fixes log de la taille à l'année  $t$  et age, et des effets aléatoires population, année et hétérogénéité individuelle.

La taille à l'année  $t + 1$  suit une distribution logNormale.

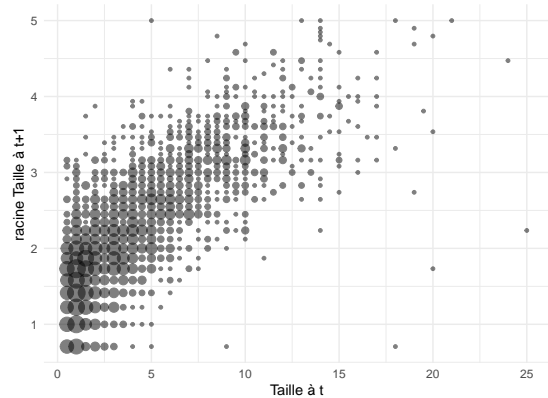
Relation entre la taille à l'année suivante et la taille à l'année actuelle



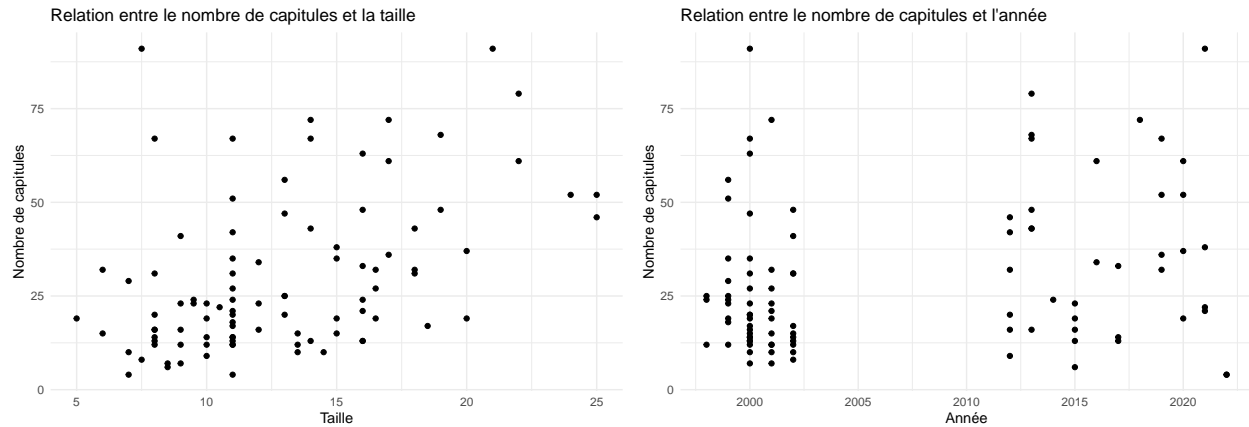
Relation log-log entre la taille à l'année suivante et la taille à l'année actuelle



Relation entre la taille à l'année suivante et la taille à l'année actuelle



## Fecondity



## Modèles par AIC

```
Survglm11 <- fitme(SurvieMars ~ 1 + poly(Size0Mars,3) +
  (Size0Mars|year) + (1|Pop),
  family=binomial,
  data=survdata1,
  method="PQL/L")
extractAIC(Survglm11)
```

```
##      edf      AIC
##    4.000 4373.984
```

```
Survglm12 <- fitme(SurvieMars ~ 1 + bs(Size0Mars,df=4,degree=2) +(bs(age0,degree=3,knots=6.5)) +
  (age0|year) + (1|Pop),
  family=binomial,
  data=survdata2,
  method="PQL/L")
extractAIC(Survglm12)
```

```
##      edf      AIC
##    9.000 2309.076
```

```
Flowglm1 <- fitme(Flowering0 ~ 1 + poly(Size0Mars,3) + poly(age0,2) +
  (age0|Pop),
  family=binomial,
  data=centauree_data, method="PQL/L")
extractAIC(Flowglm1)
```

```
##      edf      AIC
##    6.000 843.909
```

```
Pltglm1 <- fitme(Size0Mars ~ 1 + (1|year) + (1|Pop) + (1|Pop:year),
  data=plantule_data,
  family = Gamma(log))
extractAIC(Pltglm1)
```

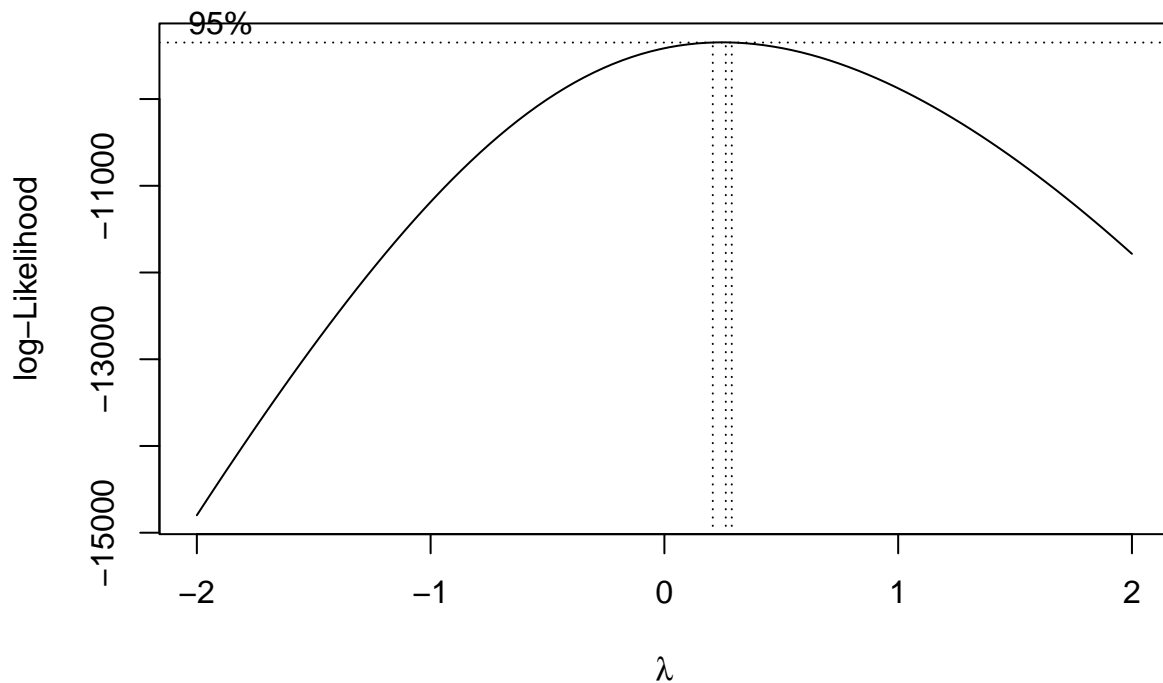
```
##      edf      AIC
##    1.000 8442.641
```

```
Growthglm1 <- fitme(log(Size1Mars) ~ 1 +
  poly(Size0Mars,4) + poly(age0,3) +
  (Size0Mars|year) + (Size0Mars|Pop),
  data=growthdata,resid.model = ~log(Size0Mars))
extractAIC(Growthglm1)
```

```
##      edf      AIC
##    8.000 3878.299
```

```
Growthglm2 <- fitme(Size1Mars ~ 1 +
  bs(Size0Mars,degree=3,df=5) + bs(age0,degree=2,knots=6.5) +
  (Size0Mars+age0|year) + (Size0Mars|Pop),
  resid.model = ~ log(Size0Mars),
  data=growthdata)

cb <- MASS::boxcox(Size1Mars ~ 1, data=growthdata)
```



```
cb$x[cb$y==max(cb$y)]
```

```
## [1] 0.2626263
```

```
Cptlglm1 <- fitme(Cptl0 ~ 1 + Size0Mars,  
                 data=cptldata)  
extractAIC(Cptlglm1)
```

```
##      edf      AIC  
## 2.0000 825.2362
```

```
library(SplinesUtils)
```

```
#Pour un exemple :
```

```
# RegModel <- fitme(SurvieMars ~ 1 + bs(Size0Mars, df = 4, degree = 2) +(bs(age0, degree = 3, knots = 6  
#  
#           family=binomial,  
#           data=survdata2,  
#           method="PQL/L")  
# SplineTerm <- "bs(Size0Mars, df = 4, degree = 2)"
```

```
# La fonction d'origine est RegSplineAsPiecePoly
```

```
SplinesAsPolySpaMM <- function (RegModel, SplineTerm, shift = TRUE) {
```

```
#Extrait les données du spline (position dans les effets fixes, degrés des polynomes, knots)  
RegSpline <- SplinesUtils:::ExtractSplineTerm(terms(RegModel), SplineTerm)  
pos <- RegSpline$pos
```

```
#Permet de sortir les coefficients de chaque bout de splines
```

```
ind0 <- attr(RegModel$X.pv, "namesOri") #Noms des coefficients  
ind <- integer(0)  
pattern <- paste0("^", gsub("\\(", "\\((", gsub("\\)", "\\))", SplineTerm)), "\\d+$")  
#C'est pas très propre mais pas le choix car les noms des bouts de splines finissent avec un indice 1  
ind <- grep(pattern, ind0) #Détection si le nom du coefficient correspond au spline voulu  
RegSplineCoef <- RegModel$fixef[ind] #Sort l'estimation des coeffs
```

```
RegSplineCoef <- unname(RegSplineCoef)
```

```
na <- is.na(RegSplineCoef)
```

```
if (any(na)) {  
  warning("NA coefficients found for SplineTerm; Replacing NA by 0")  
  RegSplineCoef[na] <- 0  
}
```

```
#Recalcule le polynome avec la méthode d'origine, je n'y ai pas touché
```

```
PiecePolyCoef <- SplinesUtils:::PiecePolyRepara(RegSpline, RegSplineCoef, shift)  
structure(list(  
  PiecePoly = list(coef = PiecePolyCoef, shift = shift),  
  knots = RegSpline$knots),  
  class = c("PiecePoly", "RegSpline"))  
}
```



```
SplinesAsPolySpaMM(Survglm12,"bs(Size0Mars, df = 4, degree = 2)")
```

```
## 3 piecewise polynomials of degree 2 are constructed!
## Use 'summary' to export all of them.
## The first 3 are printed below.
## 2.56e-16 + 1.28 * (x - 0.5) - 0.142 * (x - 0.5) ^ 2
## 2.31 + 0.569 * (x - 3) - 0.0708 * (x - 3) ^ 2
## 3.38 + 0.144 * (x - 6) - 0.00504 * (x - 6) ^ 2
```

```
SplinesAsPolySpaMM(Growthglm2,"bs(Size0Mars, degree = 3, df = 5)")
```

```
## 3 piecewise polynomials of degree 3 are constructed!
## Use 'summary' to export all of them.
## The first 3 are printed below.
## 2.19e-15 + 0.257 * (x - 0.5) + 0.609 * (x - 0.5) ^ 2 - 0.135 * (x - 0.5) ^ 3
## 1.3 + 1.17 * (x - 2) + 0.000897 * (x - 2) ^ 2 - 0.00156 * (x - 2) ^ 3
## 4.78 + 1.14 * (x - 5) - 0.0131 * (x - 5) ^ 2 - 0.00182 * (x - 5) ^ 3
```

```
SplinesAsPolySpaMM(Growthglm2,"bs(age0, degree = 2, knots = 6.5)")
```

```
## 2 piecewise polynomials of degree 2 are constructed!
## Use 'summary' to export all of them.
## The first 2 are printed below.
## 0 - 0.768 * (x - 1) + 0.104 * (x - 1) ^ 2
## -1.07 + 0.379 * (x - 6.5) - 0.661 * (x - 6.5) ^ 2
```

```
f <- function(x){
  return((2.56e-16 + 1.28 * (x - 0.5) - 0.142 * (x - 0.5) ^ 2)*I(x<3)+(2.31 + 0.569 * (x - 3) - 0.0708 * (x - 3) ^ 2)*I(x>=3))
}
```

## Détails des modèles

### Survie des plantules

$$\text{logit}(s_1(x)) = \beta_0 + b_{0,pop} + b_{0,year} + (\beta_1 + b_{1,pop})x + \beta_2 x^2 + \beta_3 x^3$$

$$\text{logit}(s_1(x)) = -1.51 + b_{0,pop} + b_{0,year} + (36.02 + b_{1,pop})x + 12.04x^2 + 29.46x^3$$

avec :

$x$  : taille au temps  $t$

$\beta_i, i \in \{0, 1, 2, 3\}$  : coefficients du polynome de la taille

$b_{0,pop} \sim \mathcal{N}(0, 0.89)$  : contribution de l'effet aléatoire Population sur l'intercept

$b_{j,year}, j \in \{0, 1\}$  : contribution de l'effet aléatoire Année sur l'intercept et la taille ( $b_{0,year}, b_{1,year} \sim \mathcal{N}(0, \Sigma)$ )  
with

$$\Sigma = \begin{bmatrix} 1.57 & -0.08 \\ -0.08 & 0.066 \end{bmatrix}$$

Correlation term:  $\rho = -0.78$

## Survie des rosettes

$$\text{logit}(s_a(x)) = s_1(x) + s_2(a)$$

$$f_1(x) = \begin{cases} b_{0,pop} + b_{0,year} + 1.28(x - 0.5) - (0.142 + b_{0,year})(x - 0.5)^2 & \text{si } x \leq 2 \\ b_{0,pop} + b_{0,year} + 2.31 + (0.569 + b_{0,year})(x - 3) - 0.0708(x - 3)^2 & \text{si } x \in [2; 6] \\ b_{0,pop} + b_{0,year} + 3.38 + (0.144 + b_{0,year})(x - 6) - 0.00504(x - 6)^2 & \text{si } x \geq 6 \end{cases}$$

$$f_2(a) = \begin{cases} b_{0,pop} + 0.612(a - 2) - 0.416(a - 2)^2 + 0.0617(a - 2)^3 & \text{si } a \leq 6.5 \\ b_{0,pop} - 0.0382 + 0.621(a - 6.5) + 0.418(a - 6.5)^2 - 0.844(a - 6.5)^3 & \text{si } a \geq 6.5 \end{cases}$$

avec :

$x$  : taille au temps  $t$

$a$  : age au temps  $t$

$b_{0,pop} \sim \mathcal{N}(0, 0.073)$  : contribution de l'effet aléatoire Population sur l'intercept

$b_{j,year}, j \in \{0, 1\}$ : contribution de l'effet aléatoire Année sur l'intercept et l'age  $(b_{0,year}, b_{1,year}) \sim \mathcal{N}(0, \Sigma)$   
with

$$\Sigma = \begin{bmatrix} 1.39 & -0.01 \\ -0.01 & 0.0097 \end{bmatrix}$$

Correlation term:  $\rho = -0.91$

## Floraison

$$\text{logit}(f_a(x)) = \beta_0 + b_{0,pop} + \beta_{1,1}x + \beta_{1,2}x^2 + \beta_{1,3}x^3 + (\beta_{2,1} + b_{1,pop})a + \beta_{2,2}a^2$$

$$\text{logit}(f_a(x)) = 11.49 + b_{0,pop} + 258.56x - 86.88x^2 + 42.73x^3 + (148.26 + b_{1,pop})a - 57.46a^2$$

avec :

$a$  : age au temps  $t$

$x$  : taille au temps  $t$

$\beta_0$  : intercept des effets fixes

$\beta_{1,i}, i \in \{1, 2, 3\}$ : coefficients du polynome de la taille

$\beta_{2,i}, i \in \{1, 2\}$ : coefficients du polynome de l'age

$b_{0,pop}$  : contribution de l'effet aléatoire Population sur l'intercept

$b_{1,pop}$  : contribution de l'effet aléatoire Population sur l'age

$(b_{0,pop}, b_{1,pop}) \sim \mathcal{N}(0, \Sigma)$  with

$$\Sigma = \begin{bmatrix} 2.17 & -0.19 \\ -0.19 & 0.09 \end{bmatrix}$$

Correlation term:

$\rho = -0.98$

## Taille des plantules

$$y \sim \text{Gamma}(\mu, \sigma)$$

avec  $\mu = \exp(\beta_0 + \gamma_0 + \varepsilon_0 + \delta_0)$  et  $\sigma = \phi \times \mu^2 = 0.32$

$y$  : taille de la plantule

$$\beta_0 = 0.21,$$

$\gamma_0 \sim \mathcal{N}(0, 0.022)$  : effet année

$\varepsilon_0 \sim \mathcal{N}(0, 0.026)$  : effet population

$\delta_0 \sim \mathcal{N}(0, 0.079)$  : effet d'interaction population:année

Residual variation :  $\phi = 0.26$

## Croissance

### log-log

$$y \sim \mathcal{N}(\mu(x, a), \sigma(x))$$

avec  $\mu(x, a) = \beta_0 + b_{0,1} + b_{0,2} + (\beta_{1,1} + b_{1,1} + b_{1,2})x + \beta_{2,1}x^2 + \beta_{3,1}x^3 + \beta_{4,1}x^4 + \beta_{2,1}a + \beta_{2,2}a^2 + \beta_{3,2}a^3$   
 $\sigma(x) = -0.9 - 0.45 \log(x)$

### 1-1

$$\mu(x, a) = \mu_1(x) + \mu_2(a)$$

$$\mu_1(x) = \begin{cases} b_{0,1,year} + b_{0,pop} + (0.257 + b_{1,1,year} + b_{1,pop})(x - 0.5) + 0.609(x - 0.5)^2 - 0.135(x - 0.5)^3 & \text{si } x \leq 2 \\ b_{0,1,year} + b_{0,pop} + 1.3 + (1.17 + b_{1,1,year} + b_{1,pop})(x - 2) + 0.000897(x - 2)^2 - 0.00156(x - 2)^3 & \text{si } x \in [2; 5] \\ b_{0,1,year} + b_{0,pop} + 4.78 + (1.14 + b_{1,1,year} + b_{1,pop})(x - 5) - 0.0131(x - 5)^2 - 0.00182(x - 5)^3 & \text{si } x \geq 5 \end{cases}$$

$$\mu_2(a) = \begin{cases} -0.768(x - 1) + 0.104(x - 1)^2 & \text{si } a \leq 6.5 \\ -1.07 + 0.379(x - 6.5) - 0.661(x - 6.5)^2 & \text{si } x \geq 6.5 \end{cases}$$

$K_{year} \sim \mathcal{N}(0, \Sigma)$  with

$$\Sigma_{year} = \begin{bmatrix} 0.34 & 0.29 & -0.74 \\ 0.29 & 0.35 & -0.42 \\ -0.74 & -0.42 & 0.042 \end{bmatrix}$$

$K_{Pop} \sim \mathcal{N}(0, \Sigma)$  with

$$\Sigma_{Pop} = \begin{bmatrix} 0.49 & -0.49 \\ -0.49 & 0.0041 \end{bmatrix}$$

## Fécondité (nombre de capitules)

$$C(x) = \beta_0 + \beta_1 x$$

$$C(x) = 2.95 + 2.07x$$

$x$  : taille au temps  $t$

$\beta_0$  : intercept

$\beta_1$ : coefficients de la taille

```

# Gyx0 <- function (y, x, a) {
#   fake <- fake_data
#   fake$age0 <- a #Fixe l'age
#   fake$Size0Mars <- unique(x) #Fixe la taille
#   sortie <- predict(Growthglm1, fake, allow.new.levels = T) #Prédit les log(taille) à t+1 possibles
#   sortie2 <- aggregate(sortie, list(fake$Size0Mars), mean) #Calcule la moyenne des log(taille)t+1
#   Sdev <- mean(residVar(Growthglm1))
#   return(dnorm(y, mean = exp(sortie2$V1), sd = exp(Sdev))) #Proba qu'un individu soit de taille y à t
# }
#
# data_predi_growth <- fake_data3 %>%
#   filter(Pop=="Au",year==2000) %>%
#   select(-year,-Pop) %>%
#   group_by(Size0Mars,age0,Size1) %>%
#   mutate(prob = Gyx0(x=Size0Mars, y=Size1,a=age0))
#
# data_predi_growth %>%
#   ggplot(aes(x = Size0Mars, y = Size1)) +
#   geom_tile(aes(fill = prob)) +
#   facet_wrap(~age0)+
#   scale_fill_viridis_c(option = "inferno",direction=-1) +
#   theme_classic()+
#   labs(title = "Croissance prédite",
#        x = "Size(t)",
#        y = "Size(t+1)",
#        fill = "Proba",
#        color = "Proba")

```

```

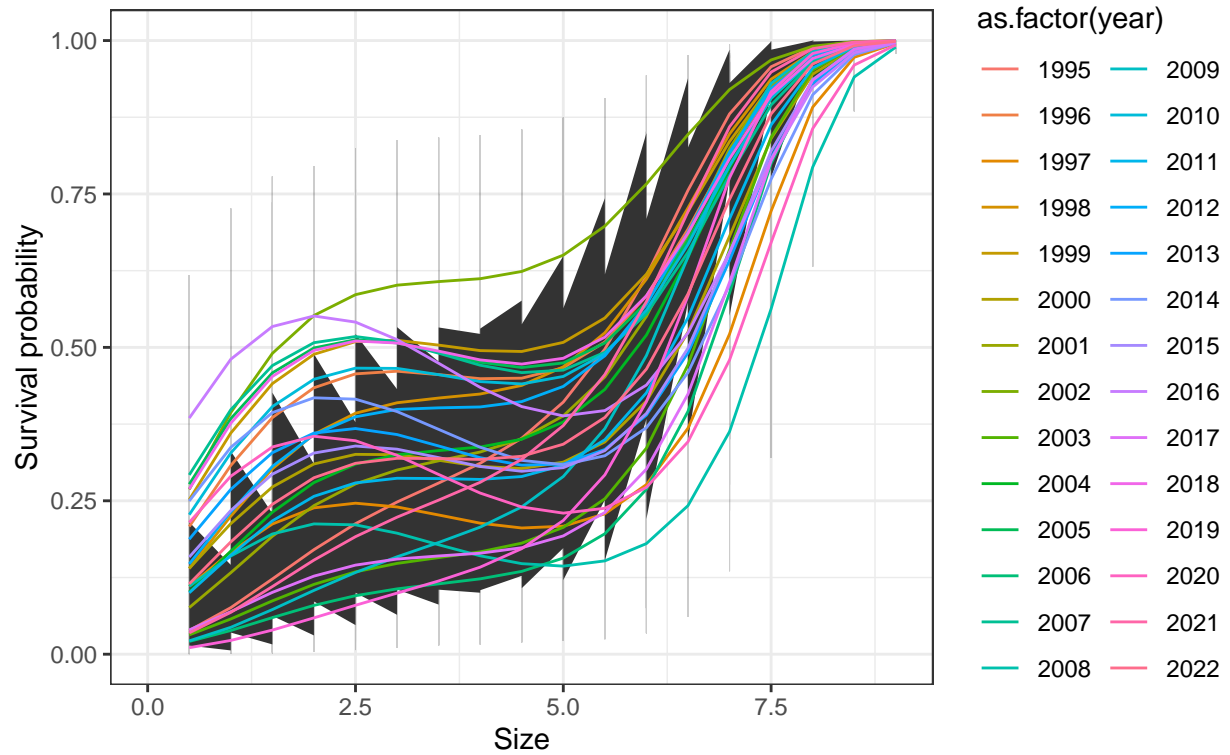
resSurv1 <- get_residVar(Survglm11,newdata = fake_data1)

fake_data1 %>%
  mutate(surv = Survpredict11,
         var = resSurv1)%>%
  group_by(Size0Mars,year) %>%
  mutate(surv_predi = mean(surv, na.rm = TRUE),
         var = mean(var)) %>%
  ggplot(aes(x = Size0Mars, y = surv_predi)) +
  geom_ribbon(aes(ymin=surv_predi-var,ymax=surv_predi+var))+
  geom_line(aes(color=as.factor(year))) +
  theme_bw()+
  ylim(0, 1) +
  xlim(0,9)+
  labs(title = "Survie prédites des plantules",
       subtitle = "Moyenne de la survie sur les années et les populations",
       x = "Size",
       y = "Survival probability")

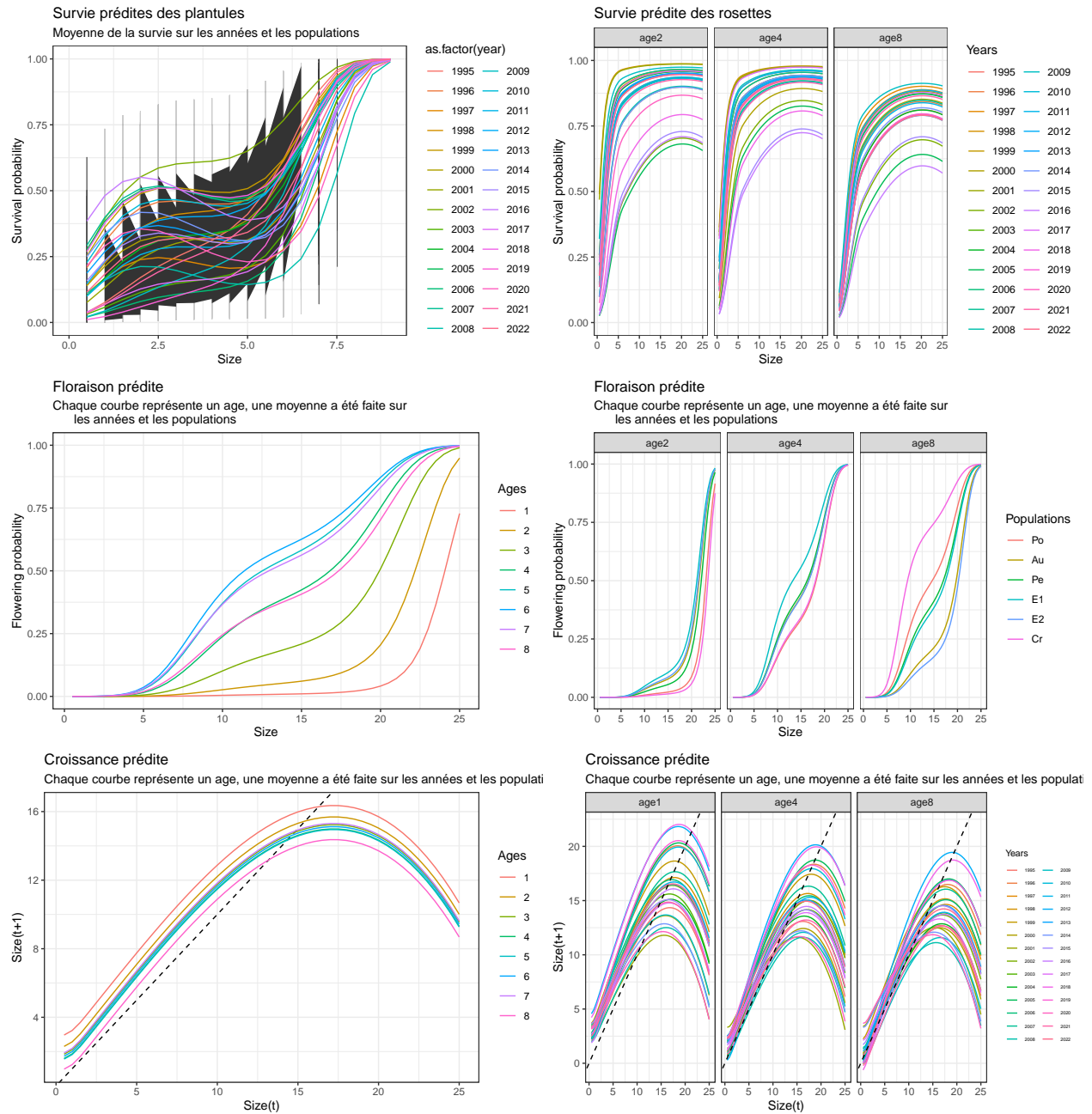
```

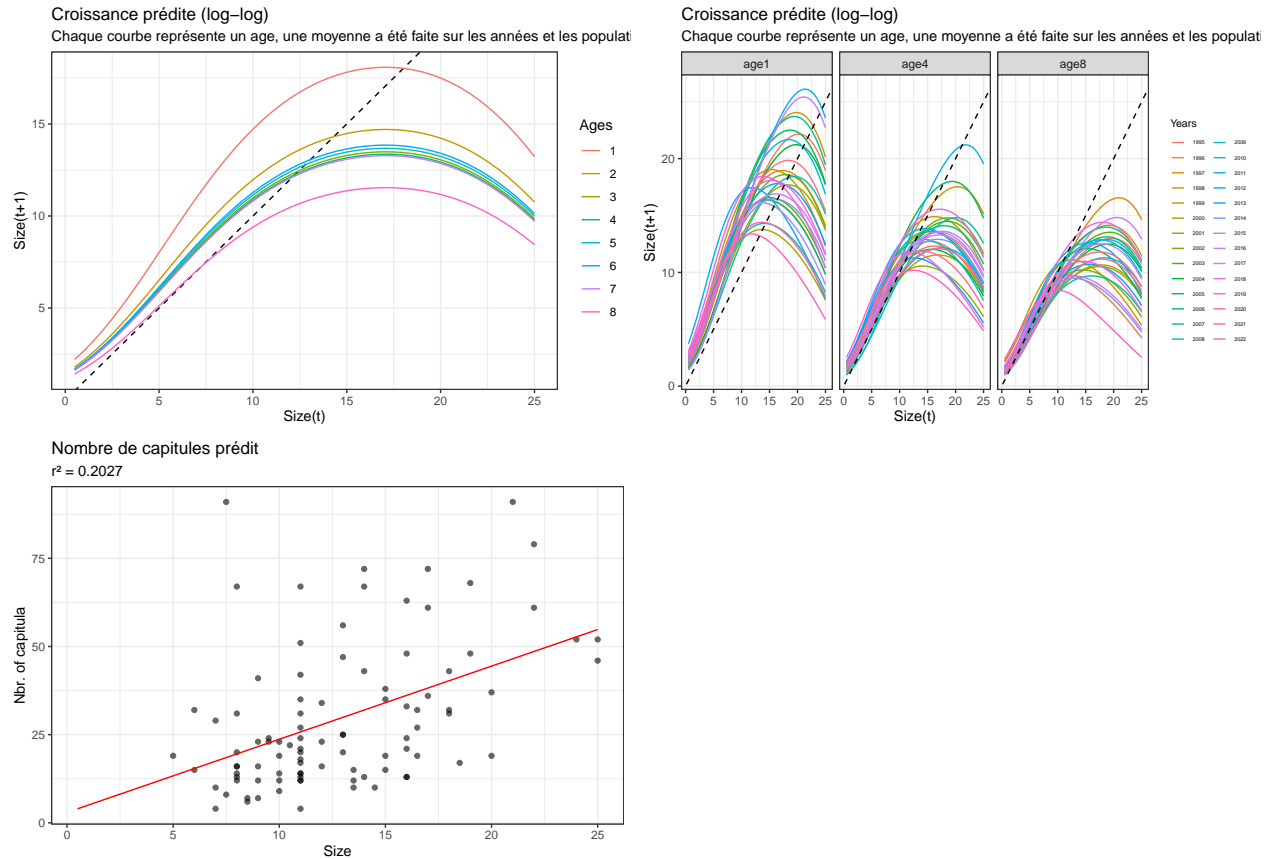
## Survie prédites des plantules

Moyenne de la survie sur les années et les populations



## Plots en fonction de la taille

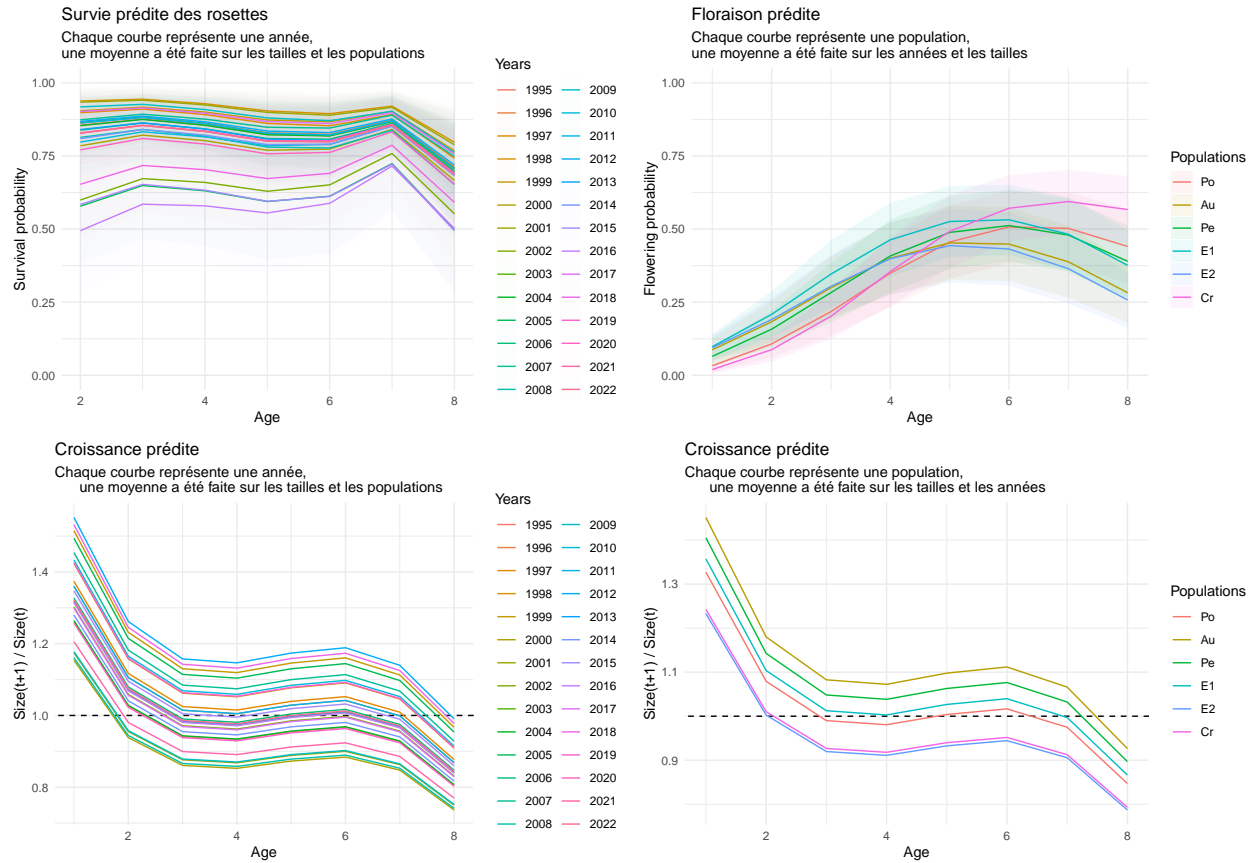




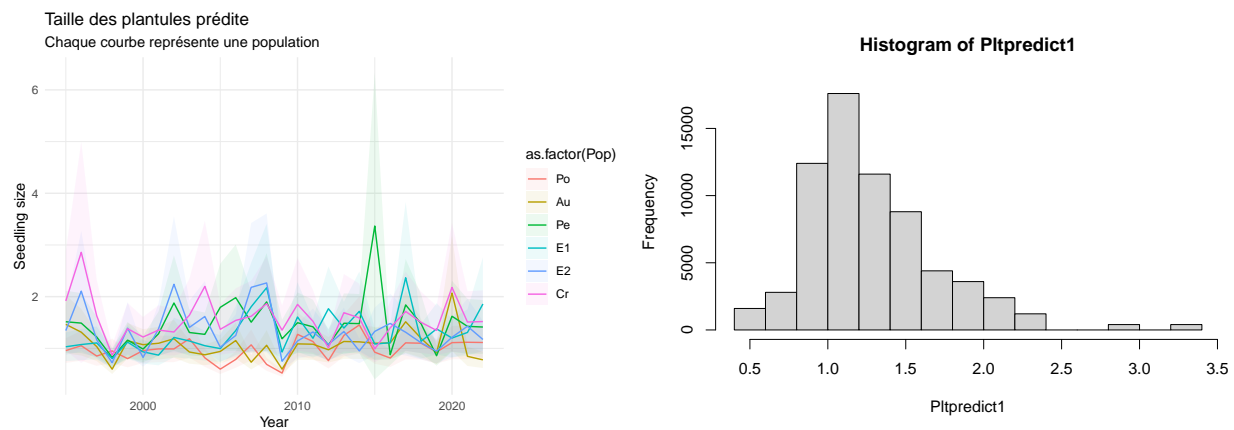
On voit clairement que la probabilité de survie augmente quand la plante grandit, jusqu'à atteindre un plateau pour les rosettes. La survie d'une plante de petite taille est assez basse, peu importe qu'elle soit d'âge 1 ou plus. On ne peut pas vraiment conclure sur la survie des plantules de grande taille ( $>5$ ) du fait de la très faible quantité de données pour ces classes de tailles (27 observations).

On voit que la probabilité de floraison augmente fortement avec la taille. De même que pour la survie, on ne peut rien dire de la floraison des plantules de grandes tailles.

## Plots en fonction de l'âge



## Autres plots



## Modèles par BIC

```
Survglm11 <- fitme(SurvieMars ~ 1+ poly(Size0Mars,3) + (1|year) + (1|Pop),
  family=binomial,
```



```

        data=survdata1,
        method="PQL/L")
extractBIC(Survglm11,ntot=length(survdata1),N=28*6)

```

```
## [1] 4390.364
```

```

Survglm12 <- fitme(SurvieMars ~ 1 + bs(Size0Mars,df=3,degree=2) + (age0|year) + (1|Pop),
        family=binomial,
        data=survdata2,
        method="PQL/L")
extractBIC(Survglm12,ntot=length(survdata2),N=28*6)

```

```
## [1] 2340.671
```

```

Flowglm1 <- fitme(Flowering0 ~ 1 + poly(Size0Mars,3) + poly(age0,2) + (age0|Pop),
        family=binomial,
        data=centauree_data, method="PQL/L")
extractBIC(Flowglm1,ntot=length(centauree_data),N=6)

```

```
## [1] 853.1186
```

```

Growthglm1 <- fitme(log(Size1Mars) ~ 1 + poly(log(Size0Mars),3) + poly(age0,3) + (log(Size0Mars)|year) +
        data=growthdata)
extractBIC(Growthglm1,ntot=length(growthdata),N=6*28)

```

```
## [1] 4113.802
```

```

Pltglm1 <- fitme(Size0Mars ~ 1 + (1|year) + (1|Pop) + (1|Pop:year),
        data=plantule_data,
        family = Gamma(log))
extractBIC(Pltglm1, ntot = length(plantule_data), N=28*6*147)

```

```
## [1] 8473.623
```

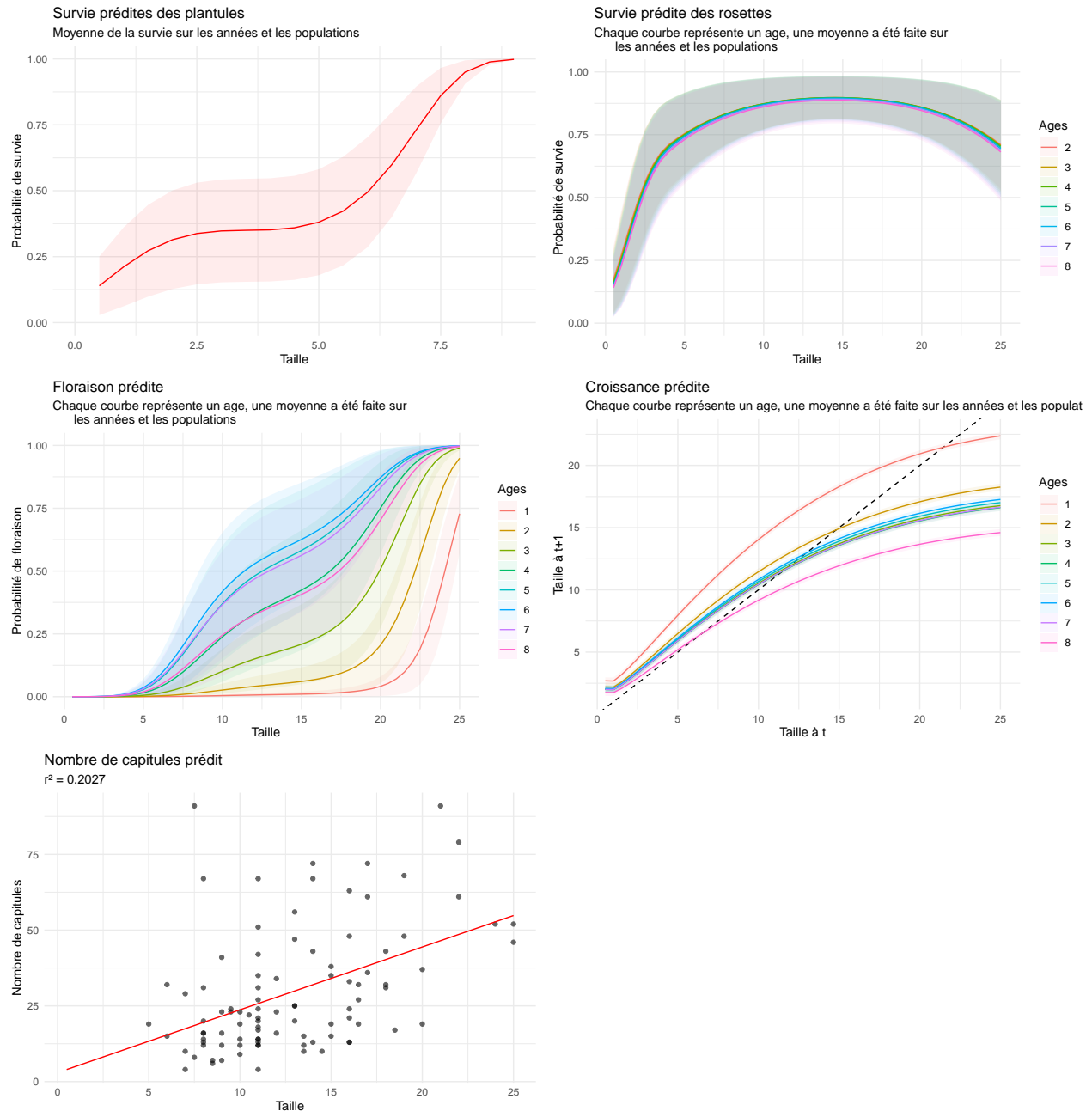
```

Cptlglm1 <- fitme(Cptl0 ~ 1 + Size0Mars,
        data = cptldata)
extractBIC(Cptlglm1,ntot = length(cptldata),N=1)

```

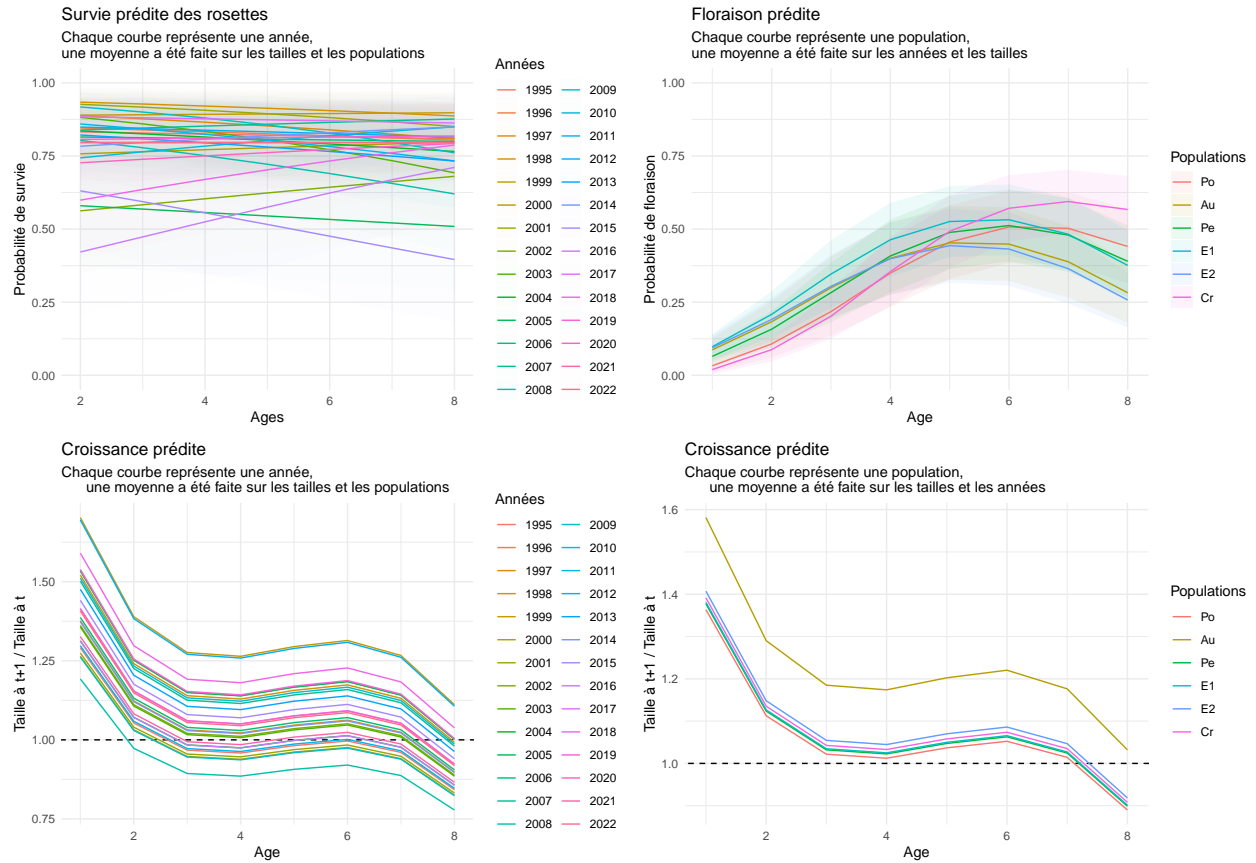
```
## [1] 826.5143
```

## Plots en fonction de la taille



Les modèles choisis par BIC sont globalement similaires que pour ceux choisis par AIC. On constate quand même des différences pour la survie des rosettes et pour la croissance. Pour la survie des rosettes, le modèle choisi propose une diminution de la survie pour les plantes de grande taille. Cette incertitude avec le modèle AIC correspond avec l'incertitude qu'on a pour ces classes de taille qui comportent peu de données. Pour la croissance, l'effet de réduction de la croissance pour les plantes de grande taille se retrouve dans les deux modèles (AIC et BIC), cependant cet effet est moins fort pour le modèle BIC.

## Plots en fonction de l'âge



La sélection par le BIC réduit le nombre de variables explicatives. On peut le voir notamment pour la survie des rosettes en fonction de l'âge. Le modèle BIC ne choisit pas une fonction de survie qui dépend de l'âge, il y a seulement un effet aléatoire de l'âge sur l'année.

## Autres plots

