

# Projet HAB904B

Basile Pajot (DARWIN), Marion Themeze-Leroy (ECOSYSTEMES),

2023-12-21

## Contents

<b>1</b>	<b>Lecture et exploration des données</b>	<b>1</b>
1.1	La variable à expliquer . . . . .	1
1.2	Les variables explicatives . . . . .	1
1.3	Exploration des données . . . . .	2
<b>2</b>	<b>Ajustement d'un modèle simple</b>	<b>5</b>
<b>3</b>	<b>Comparaison de modèles</b>	<b>9</b>
<b>4</b>	<b>Inférence et interprétation des résultats</b>	<b>38</b>
<b>5</b>	<b>Discussion</b>	<b>38</b>

## 1 Lecture et exploration des données

### 1.1 La variable à expliquer

La variable d'intérêt à expliquer, est **Shells**, soit le nombre de carapaces de tortues récentes trouvées lors des relevés sur le terrain. Cette variable est un proxy pour estimer le nombre de tortues mortes d'une année sur l'autre.

### 1.2 Les variables explicatives

- **Prev** est une variable explicative qualitative qui correspond à la prévalence pour *Mycoplasma agassizii*, soit le rapport entre le nombre de tortues séropositives sur l'effectif total de tortues par année pour chaque site.
- **Site** est une variable qualitative qui correspond au site d'échantillonnage. Elle a 10 modalités : le parc national *Big Shoals (BS)*, l'aire de gestion de la faune sauvage *Camp Blanding (CB)*, l'aire de gestion de la faune sauvage et de l'environnement *Cecil Field/Branan Field (CF)*, une *propriété privé en Floride centrale (CE)*, le parc national *Fort Cooper (FC)*, l'aire de gestion de la faune sauvage *Flying Eagle (FE)*, le parc national *Gold Head Branch (GH)*, l'aire de gestion de la faune sauvage et de l'environnement *Perry Oldenburg (OL)*, la station biologique *Ordway-Swisher (OR)*, l'aire de gestion de la pêche *Tenoroc Fish (TE)*.
- **Area** est une variable quantitative qui correspond à l'aire couverte par site lors des relevés.
- **Year** est une variable qualitative qui correspond à l'année pour laquelle les relevés ont été faits. Elle a 3 modalités : 2004, 2005, 2006.

### 1.3 Exploration des données

Nous regardons un résumé statistique des variables de notre jeu de donnée.

```
##      Site      year      shells      type      Area
## BS      : 3   Length:30   Min.    :0.00   Fresh:30   Min.    : 5.30
## CB      : 3   Class :character 1st Qu.:0.00           1st Qu.:15.20
## Cent    : 3   Mode  :character Median :1.00           Median :27.30
## CF      : 3                      Mean  :1.80           Mean   :29.02
## FC      : 3                      3rd Qu.:2.75          3rd Qu.:43.20
## FE      : 3                      Max.    :9.00           Max.    :61.00
## (Other):12
##      density      prev      total_turtle      standprev
## Min.    : 1.80   Min.    : 1.00   Min.    : 44.80   Min.    : -0.9163
## 1st Qu.: 2.50   1st Qu.: 1.20   1st Qu.: 74.75   1st Qu.: -0.9088
## Median : 3.50   Median :15.95   Median :104.22   Median : -0.3559
## Mean    : 8.76   Mean    :25.45   Mean    :119.72   Mean    : 0.0000
## 3rd Qu.: 4.80   3rd Qu.:42.05   3rd Qu.:174.90   3rd Qu.: 0.6223
## Max.    :33.00   Max.    :80.70   Max.    :200.79   Max.    : 2.0709
##
##      H      Cov_2004      Cov_2005      Cov_2006
## Min.    :0.0000   Min.    :0.0000   Min.    :0.0000   Min.    :0.0000
## 1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
## Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000
## Mean    :0.4333   Mean    :0.3333   Mean    :0.3333   Mean    :0.3333
## 3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
## Max.    :1.0000   Max.    :1.0000   Max.    :1.0000   Max.    :1.0000
##
```

Nous avons un plan d'expérience équilibré avec un même nombre d'observations par site et par année.

Pour la prevalence `prev` la moyenne est supérieure à la médiane, c'est-à-dire que plus de 50% des valeurs sont inférieures à la moyenne. Il en est de même pour le nombre de carapces `shells`. De plus, pour la prévalence, la différence entre le troisième quartile et le minimum est d'environ 40, tout comme la différence entre le maximum et le 3ème quartile. Ainsi, la gamme de valeurs prise par 25% des données est égales à celle prise par 75% des données. Pour le nombre de carapaces, la gamme de valeurs prise par 25% des données est plus de trois fois supérieur à celle prise par 75% des données. Ceci est illustré par les *figures 1 et 2*.

```
## NULL
```

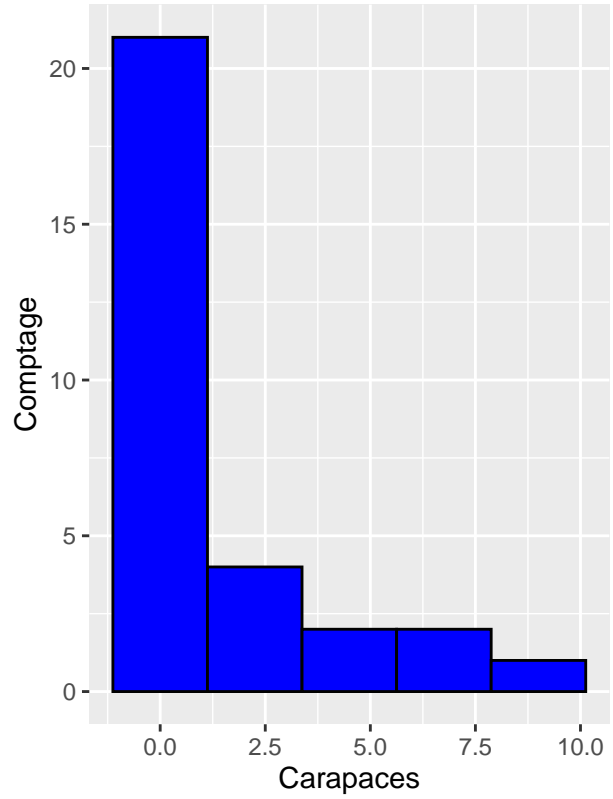
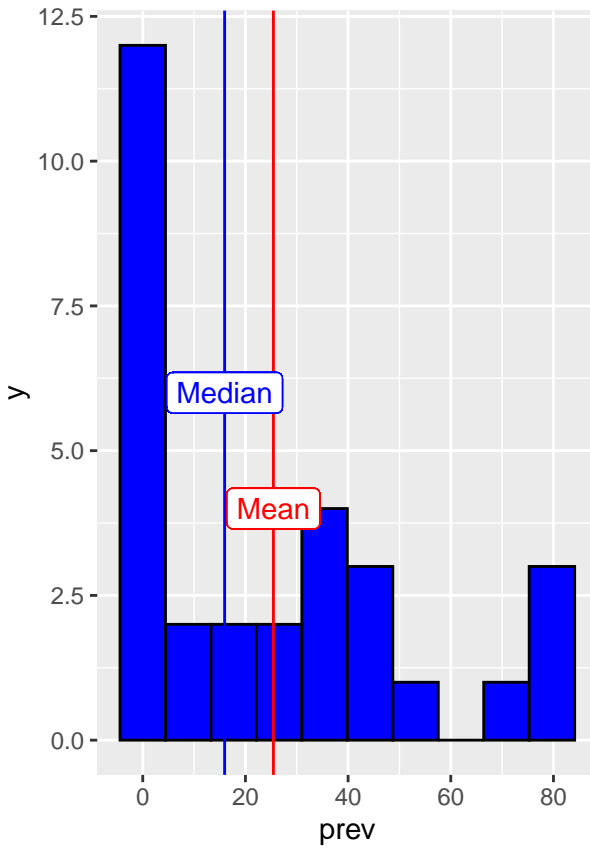
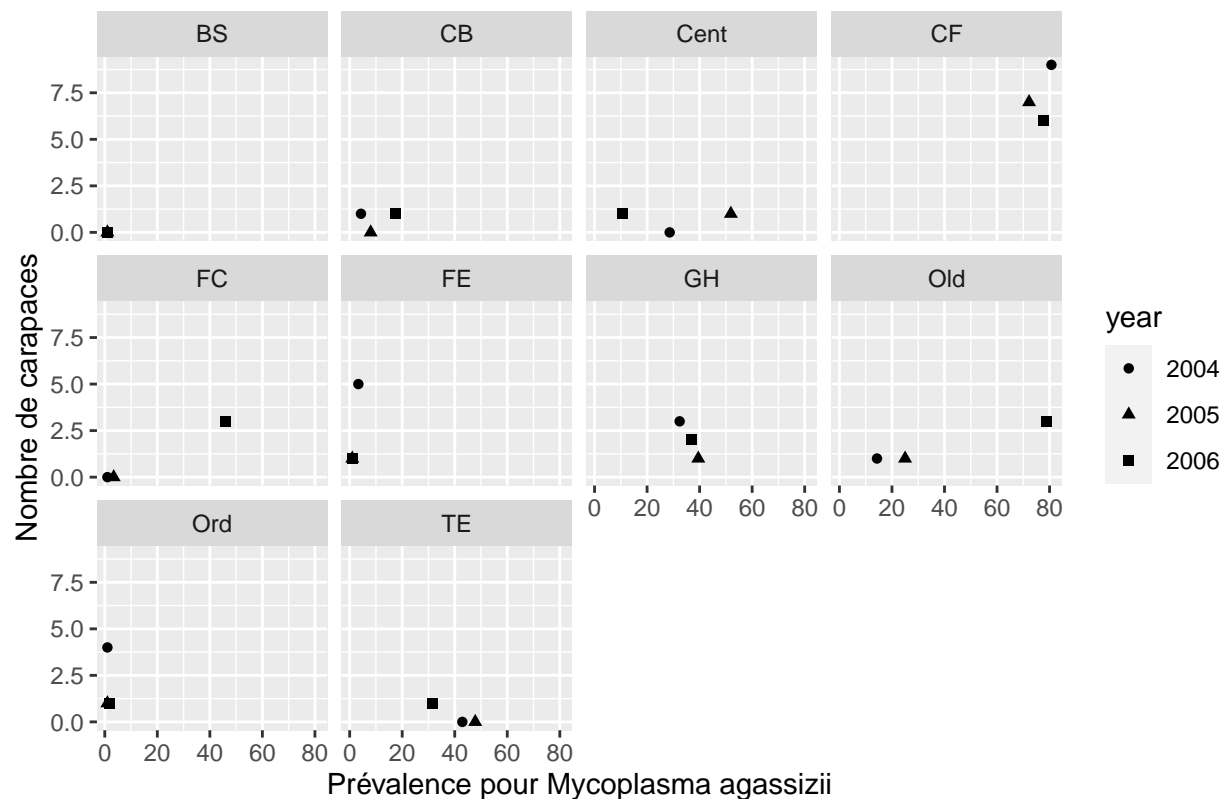


Figure 2 : Distribution du nombre de  $c$

La distribution du nombre de carapaces ressemble à une distribution de Poisson.

La figure 3 donne plusieurs informations sur notre jeu de données.



*Figure 3 : Le nombre de carapaces en fonction de la prevalence par site par année*

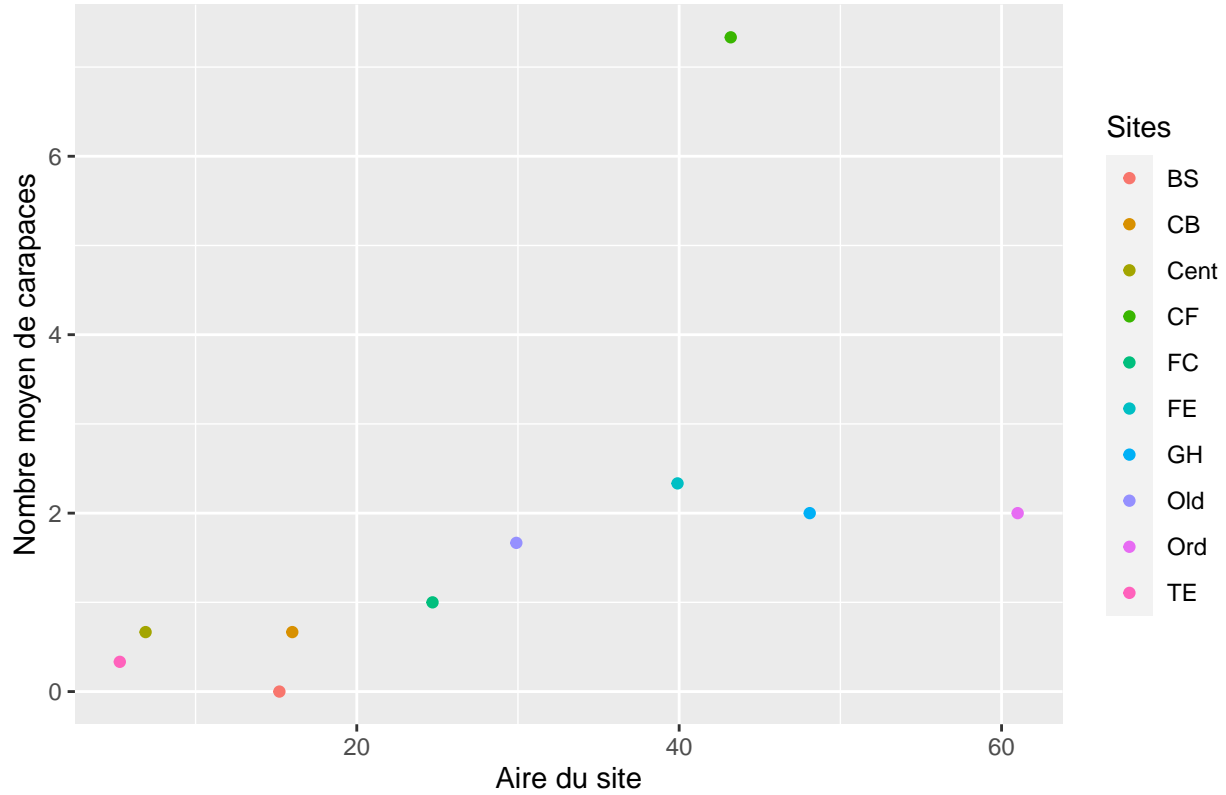
Tout d'abord, le nombre de carapaces récentes trouvées varie ou non en fonction des années et cette variation n'est pas la même en fonction des sites. Le nombre de carapaces récentes par rapport à l'année précédente reste constante augmente ou diminue. On observe des évolutions différentes pour les sites : on observe une diminution du nombre de carapaces pour le site CF sur les trois ans ou un changement de tendance se traduisant par une diminution puis une augmentation pour le site CB.

Ensuite, la prévalence en fonction des sites peut également varier en fonction des années. Comme précédemment, cette variation n'est pas la même en fonction des sites. La prévalence reste constante, augmente ou diminue. La variation peut être globale sur les trois années d'étude (augmentation de la prévalence pour le site Old) ou changer (augmentation puis diminution pour le site Cent). Lorsque la prevalence augmente d'une année à l'autre  $prev[n] < prev[n+1]$ , le nombre de carapaces récentes trouvées l'année suivante augmente  $shells[n+2] > shells[n+1]$  (sites CB, Old), et inversement (sites CF).

Ainsi, les variations du nombre de carapaces récentes trouvées pourrait être expliquée par les variations de la prévalence.

Il apparaît également que certains sites ont de faibles prévalences (BS, Ord, FE) quelque soit l'année et d'autres des prévalences élevées (CF, GH, TE). Ceci concourt avec les observations faites précédemment avec le résumé de la variable `prev` et la distribution de la prévalence. Ainsi, nous pourrions séparer les sites en deux catégories, ceux à faible ou forte prévalence. Cette variables sera donc traitée de deux manières : de manière continue et de manière discontinue avec deux catégories : - faible prévalence (0) :  $Prev < 0.25$  - haute prévalence (1) :  $Prev > 0.25$

Nous allons donc essayer de déterminer si la prévalence et l'année permettent d'expliquer les variations du nombre de carapaces. Nous avons vu que la prévalence et le nombre de carapaces trouvées diffère entre les sites et entre les années. Afin de pouvoir nous concentrer sur l'effet de la prévalence, nous allons mettre un effet aléatoire sur la variable `Site`. Nous pourrions faire de même pour la variable `Année` mais par souci de simplification, nous allons garder cette variable en effet fixe.



*Figure 4 : Le nombre moyen de carapaces en fonction de l'aire du site d'étude*

D'après la *figure 4* les sites n'ont pas tous la même aire et il semble qu'un plus grand nombre de carapces sont trouvés sur les sites avec une plus grande aire. Afin de pourvoir comparer les sites entre eux, nous allons prendre le rapport entre le nombre de carapaces trouvées par site et l'aire du site.

D'après les observations faites précédemment, nous souhaitons donc déterminer si : - le nombre de carapaces récentes trouvées est corrélée avec la prevalence pour *Mycoplasma agassizii* pour une année donnée. - le nombre de carapces récentes trouvées est plus grand dans les sites à haute prévalence par rapport aux sites à basse prévalence.

## 2 Ajustement d'un modèle simple

Nous commençons par un modèle simple M1 en considérant uniquement l'effet de la prévalence sur le nombre de carapces récentes trouvées.

L'équation (approche fréquentiste) du modèle linéaire simple est la manière suivante :

$$\frac{shells}{aire_{site}} = \mu_0 + \beta * prev$$

Ceci se traduit en approche bayésienne par un modèle considérant les hypothèses suivantes :

- **shells** suit un loi de poisson de paramètre  $\lambda$  (*cf figure 2*), c'est-à-dire que c'est une variable discrète de comptage dans un intervalle de temps et un espace donnés ; avec une variance égale à la moyenne  $E(shells) = V(shells) = \lambda$
- toutes les observations de **shells** sont **indépendantes**
- le logarithme de la moyenne de **shells** peut être exprimée comme la combinaison linéaire des variables explicatives sélectionnées.
- que les paramètres à estimer (ordonnée à l'origine et coefficients de regression) suivent des lois connues, explicités ci-après.

Nous avons donc :

$Shells_i \stackrel{i.i.d}{\sim} Pois(\lambda_i)$  avec  $i = 1, \dots, 30$  le nombre d'observations

$$\log(\lambda_i) = \log(aire_i) + \mu_0 + \beta * prev_i$$

Nous utilisons comme priors les distribution suivantes :

$$\mu_0 \sim \mathcal{N}(0, 100)$$

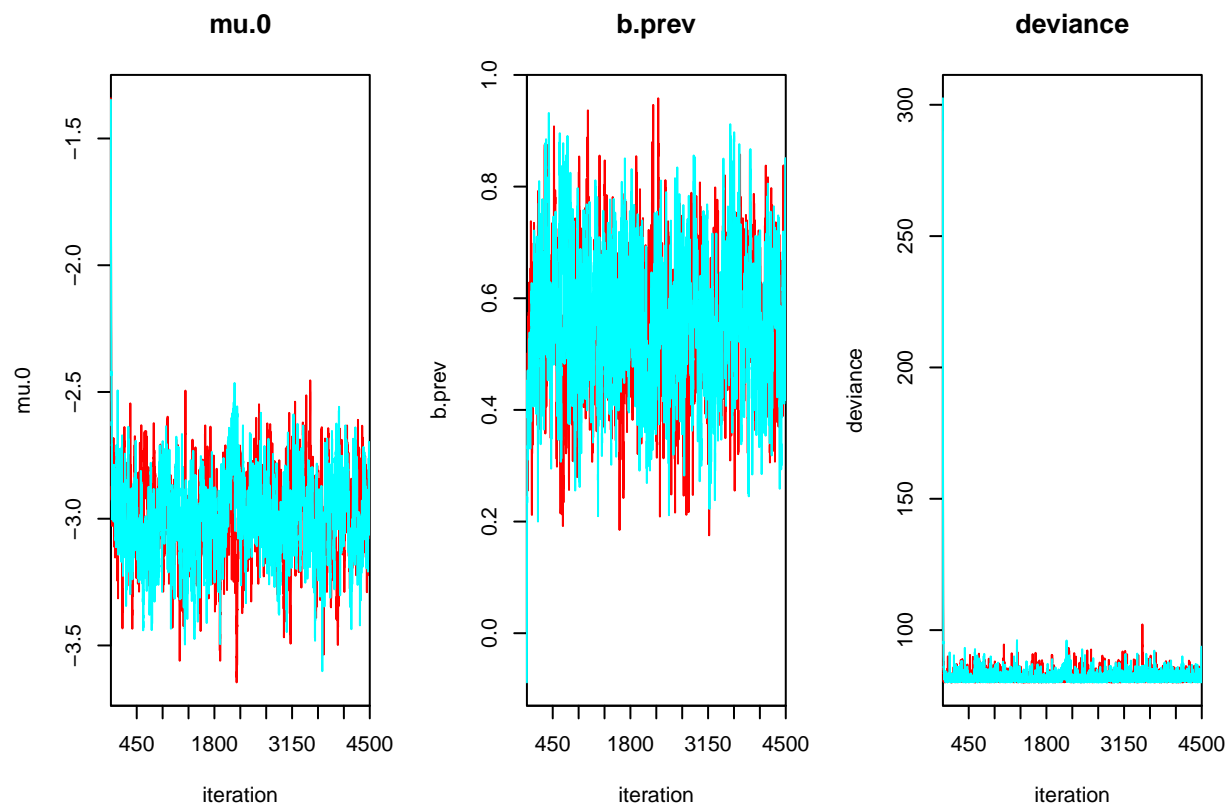
$$\beta \sim \mathcal{N}(0, 100)$$

```
## module glm loaded

## Compiling model graph
##   Resolving undeclared variables
##   Allocating nodes
## Graph information:
##   Observed stochastic nodes: 30
##   Unobserved stochastic nodes: 2
##   Total graph size: 182
##
## Initializing model

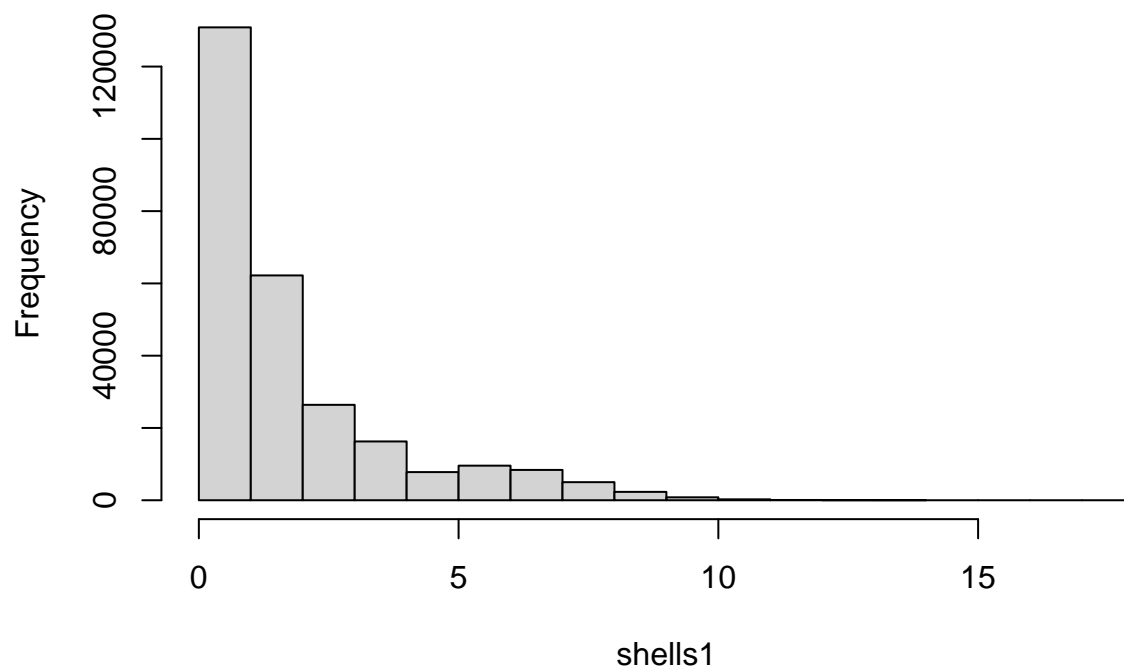
## Inference for Bugs model at "C:/Users/basil/AppData/Local/Temp/RtmpaCcf2h/model79e83ea64b8f.txt", fi
## 2 chains, each with 9000 iterations (first 4500 discarded)
## n.sims = 9000 iterations saved
##           mu.vect sd.vect   2.5%   25%   50%   75%  97.5%  Rhat n.eff
## b.prev      0.554   0.116  0.334  0.474  0.553  0.633  0.785 1.002  1200
## mu.0       -2.993   0.168 -3.321 -3.103 -2.991 -2.884 -2.678 1.001  6100
## deviance   82.236   4.363 80.155 80.685 81.487 82.848 87.763 1.001  9000
##
## For each parameter, n.eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor (at convergence, Rhat=1).
##
## DIC info (using the rule, pD = var(deviance)/2)
## pD = 9.5 and DIC = 91.8
## DIC is an estimate of expected predictive error (lower deviance is better).
```

Nous vérifions que le modèle a bien convergé.



Les 2 chaînes se mélangent bien et convergent toutes deux. Ceci est aussi confirmé par la statistique de Gelman-Rubin  $\hat{R}$  qui est inférieure à 1.1 pour chaque paramètre estimé. Nous notons aussi que `n.eff` est supérieur à 100. Nous avons donc des estimations de nos paramètres qui sont stables et des chaînes peu autocorrélées.

## Histogram of shells1



```
## [1] 1.815418
```

```
shells1 <- shells1 %>% as.matrix(ncol=1) %>% as.data.frame() %>% rename(shells=x)
```



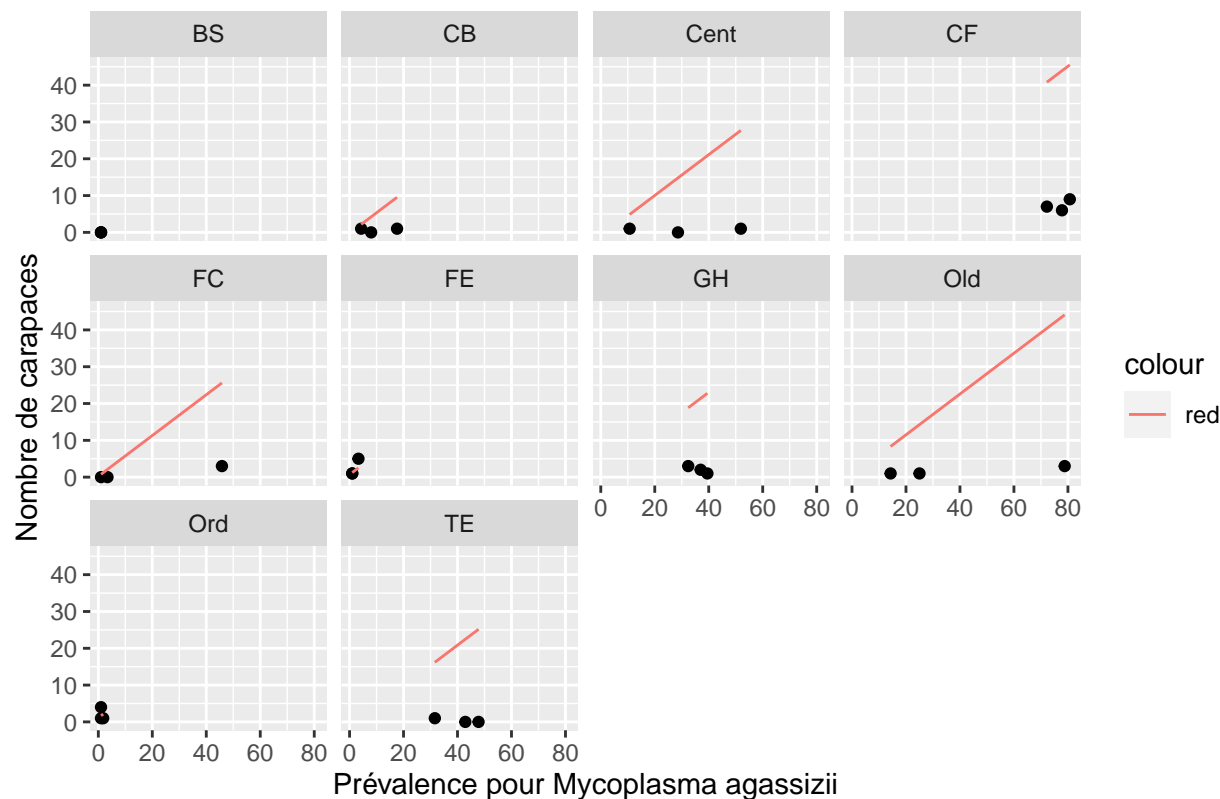


Figure 5 : Le nombre de carapaces en fonction de la prevalence par site par année

Notre modèle M1 prédit bien la distribution des carapaces mais elle ne prend pas en compte la variabilité des sites comme le montre les droites de regressions construites à partir de la moyenne de distributions postérieures de nos paramètres (cf. figure 5).

Comme vu dans la partie 1, le nombre de carapace varie en fonction de l'année et du site. Nous allons donc ajuster différents modèles qui prennent en compte ces deux variables individuellement ou en les combinant.

### 3 Comparaison de modèles

Une analyse statistique passe toujours par l'ajustement d'un premier modèle simple. Commencez par décrire le modèle qui vous semble pertinent pour répondre à une des questions traitées dans l'article, en vous aidant d'équations ou pas, et en justifiant son utilisation. Rappelez les hypothèses sur lesquelles repose le modèle. Notez qu'il ne vous est pas demandé de vérifier la validité de ces hypothèses (l'analyse des résidus). Vous pouvez ignorer les effets aléatoires sur les pentes (slopes) et ne considérer que des modèles avec un effet aléatoire sur l'ordonnée à l'origine (intercept).

En vous aidant de l'article, formez quelques hypothèses et construisez les modèles correspondant. Ajustez et comparez ces modèles pour déterminer l'hypothèse la mieux supportée par les données. N'oubliez de standardiser les variables explicatives continues avec la fonction `scale()` de R par exemple. Dans le cas d'une variable année, la standardisation est un peu différente, n'utilisez pas la fonction `scale()`. Si `year <- c(2006, 2007, ... 2021)` par exemple, utilisez simplement la variable `year - 2005` qui vaut (1, 2, ... 16) pour avoir toujours une variable entière, mais qui commence à 1 et ne prend plus de grandes valeurs.

Le modèle null M0 :

$$\log(\lambda_i) = \log(\text{aire}_i) + \mu_0$$

```
# The null model function
null_model <- function(){
```

```

# The null model does not take into account any of the variables. We just want to see the evolution of
# Likelihood
for (i in 1:N){ # Loop over observations
  S[i] ~ dpois(lambda[i])
  log(lambda[i]) <- mu.0 + log(A[i])
}
# Priors
mu.0 ~ dnorm(0, 1/100)
}

# Now, we can make list of the data to use in the jags function
datax <- list(
  N = gopher$year %>%
    length(),
  S = gopher$shells,      # Number of shells
  A = gopher$Area        # Area offset
)

# Parameters to estimate
params <- c("mu.0")

# Initialising the chains
init1 <- list(mu.0 = -0.5)
init2 <- list(mu.0 = 0.5)
init <- list(init1, init2)

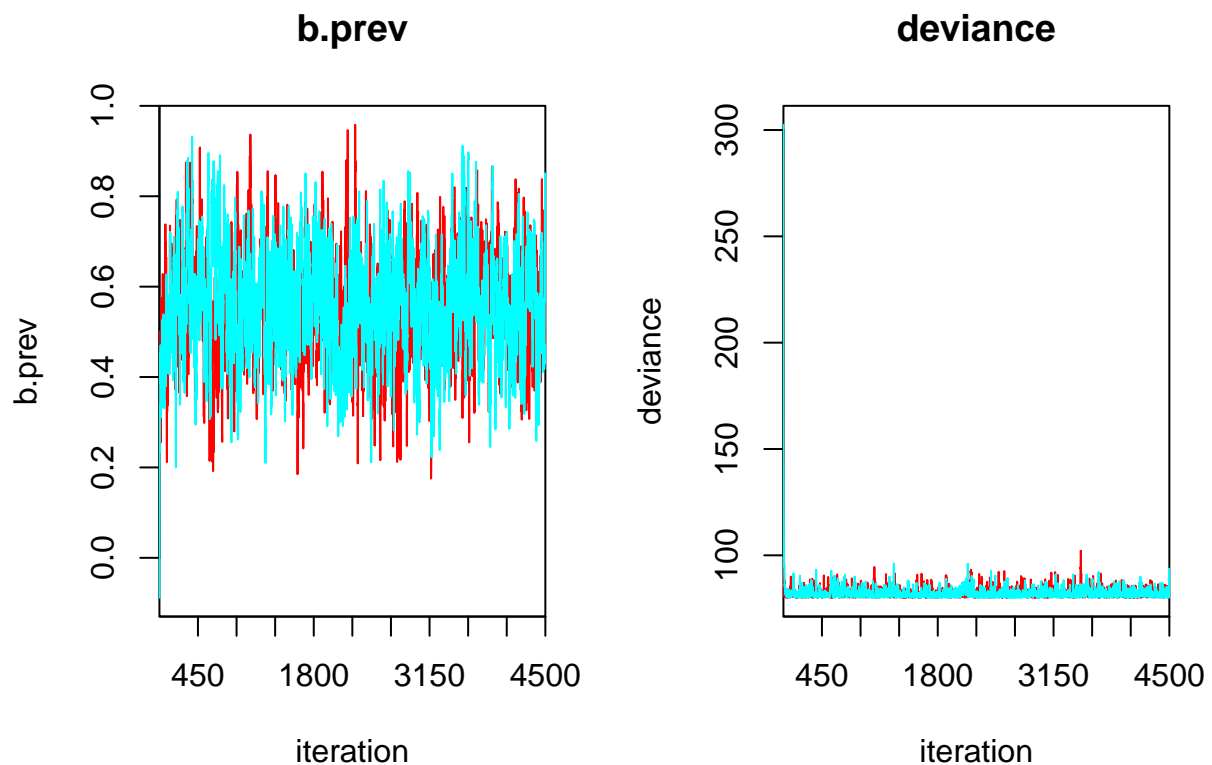
# Define the iteration parameters
nb.iterations <- 9000
nb.burnin <- 4500
# Run the model using jags
M0 <- jags(
  data = datax,
  inits = init,
  parameters.to.save = params,
  model.file = null_model,
  n.chains = 2,
  n.iter = nb.iterations,
  n.burnin = nb.burnin,
  n.thin = 1
)

## Compiling model graph
##   Resolving undeclared variables
##   Allocating nodes
## Graph information:
##   Observed stochastic nodes: 30
##   Unobserved stochastic nodes: 1
##   Total graph size: 96
##
## Initializing model
M0

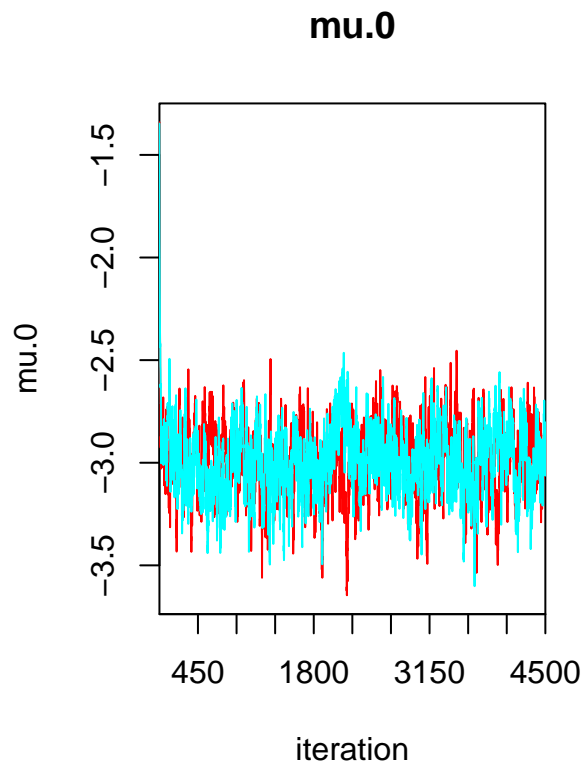
## Inference for Bugs model at "C:/Users/basil/AppData/Local/Temp/RtmpaCcf2h/model179e84857db3.txt", fit
## 2 chains, each with 9000 iterations (first 4500 discarded)

```

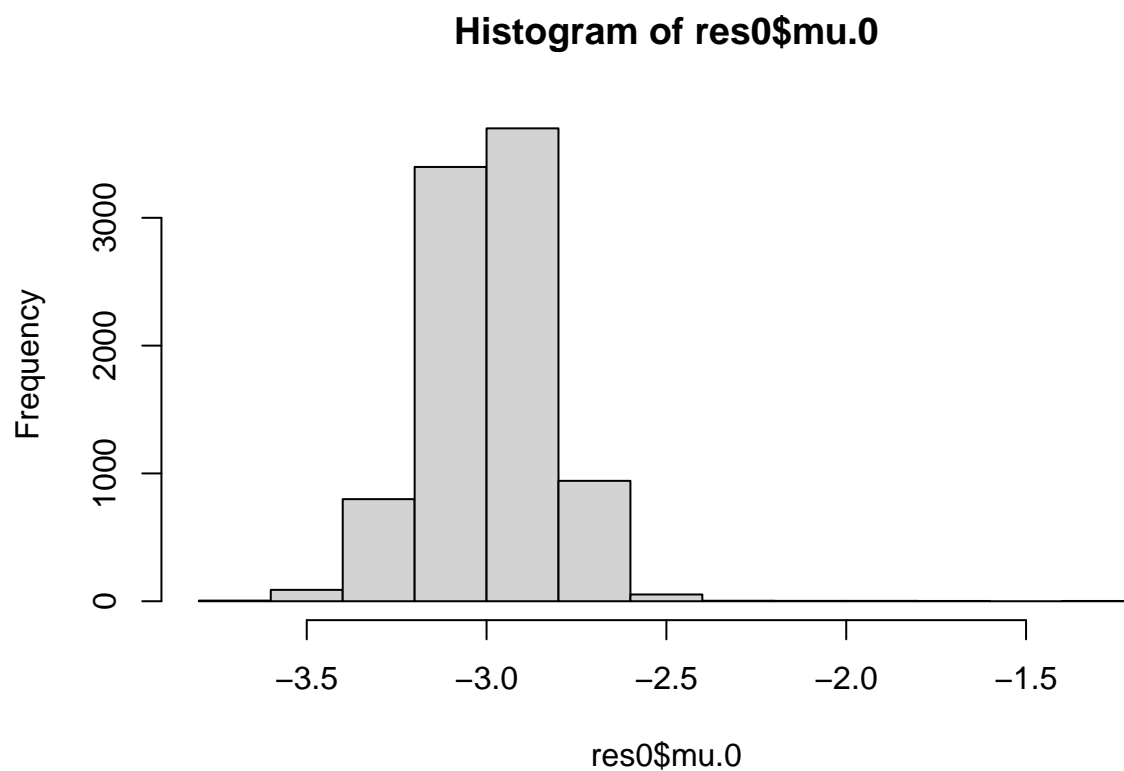
```
## n.sims = 9000 iterations saved
##      mu.vect sd.vect  2.5%    25%    50%    75%   97.5% Rhat n.eff
## mu.0      -2.787  0.137  -3.069  -2.879  -2.782  -2.692  -2.530 1.001  9000
## deviance 104.666  1.406 103.656 103.759 104.130 105.010 108.707 1.001  9000
##
## For each parameter, n.eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor (at convergence, Rhat=1).
##
## DIC info (using the rule, pD = var(deviance)/2)
## pD = 1.0 and DIC = 105.7
## DIC is an estimate of expected predictive error (lower deviance is better).
# Evaluons la convergence du modèle
traceplot(M1, mfrow=c(1, 2), ask=FALSE)
```



```
par(mfrow = c(1, 1))
```

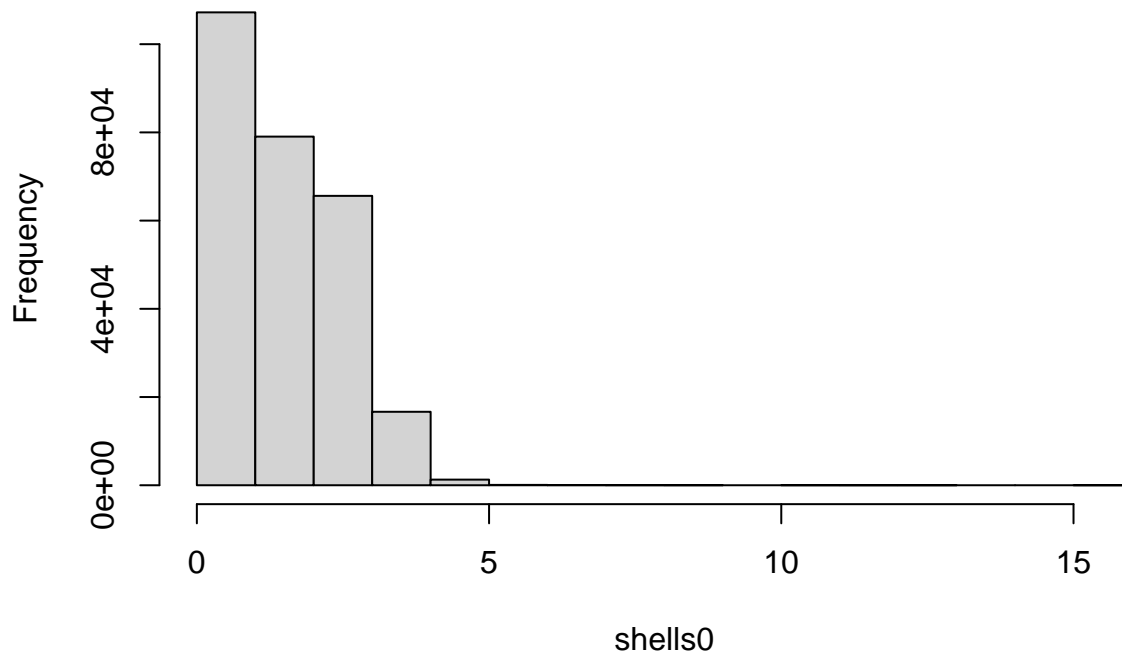


```
# On récupère les paramètres et on regarde leur distribution  
res0 <- M1$BUGSoutput$sims.matrix %>%  
  as.data.frame()  
hist(res0$mu.0)
```



```
# Rétrotransformation
shells0 <- matrix(NA, ncol=nrow(gopher), nrow=nrow(res0))
for (i in 1:nrow(gopher)){
  shells0[,i] <- gopher$Area[i] * exp(res0$mu.0)
}
hist(shells0)
```

## Histogram of shells0



```
mean(shells0)
```

```
## [1] 1.476898
```

```
# Récupérons le DIC
```

```
DICO <- M0$BUGSoutput$DIC
```

Un modèle avec uniquement l'année M3 :

$$\log(\lambda_i) = \log(\text{aire}_i) + \mu_0 + \alpha_0 * \text{cov}_{2004} + \alpha_1 * \text{cov}_{2005} + \alpha_2 * \text{cov}_{2006}$$

```
# The year model function
```

```
year_model <- function(){
```

```
  # This model takes into account the variable year with it's 3 modalities.
```

```
  # Likelihood
```

```
  for (i in 1:N){ # Loop over observations
```

```
    S[i] ~ dpois(lambda[i])
```

```
    log(lambda[i]) <- mu.0 + alpha_0 * Cov_2004[i] + alpha_1 * Cov_2005[i] + alpha_2 * Cov_2006[i] + log
```

```
  }
```

```
  # Priors
```

```
  mu.0 ~ dnorm(0, 1/100)
```

```
  alpha_0 ~ dnorm(0, 1/100)
```

```
  alpha_1 ~ dnorm(0, 1/100)
```

```
  alpha_2 ~ dnorm(0, 1/100)
```

```
}
```

```
# Now, we can make list of the data to use in the jags function
```

```
datax <- list(
```

```

N = gopher$year %>%
  length(),
S = gopher$shells,      # Number of shells
A = gopher$Area,        # Area offset
Cov_2004 = gopher$Cov_2004, # Effect of the year 2004
Cov_2005 = gopher$Cov_2005, # Effect of the year 2005
Cov_2006 = gopher$Cov_2006 # Effect of the year 2006
)

# Parameters to estimate
params <- c("mu.0", "alpha_0", "alpha_1", "alpha_2")

# Initialising the chains
init1 <- list(mu.0 = -0.5, alpha_0=-0.5, alpha_1=-0.5, alpha_2=-0.5)
init2 <- list(mu.0 = 0.5, alpha_0=0.5, alpha_1=0.5, alpha_2=0.5)
init <- list(init1, init2)

# Define the iteration parameters
nb.iterations <- 9000
nb.burnin <- 4500
# Run the model using jags
M3 <- jags(
  data = datax,
  inits = init,
  parameters.to.save = params,
  model.file = year_model,
  n.chains = 2,
  n.iter = nb.iterations,
  n.burnin = nb.burnin,
  n.thin = 1
)

```

```

## Compiling model graph
##   Resolving undeclared variables
##   Allocating nodes
## Graph information:
##   Observed stochastic nodes: 30
##   Unobserved stochastic nodes: 4
##   Total graph size: 235
##
## Initializing model

```

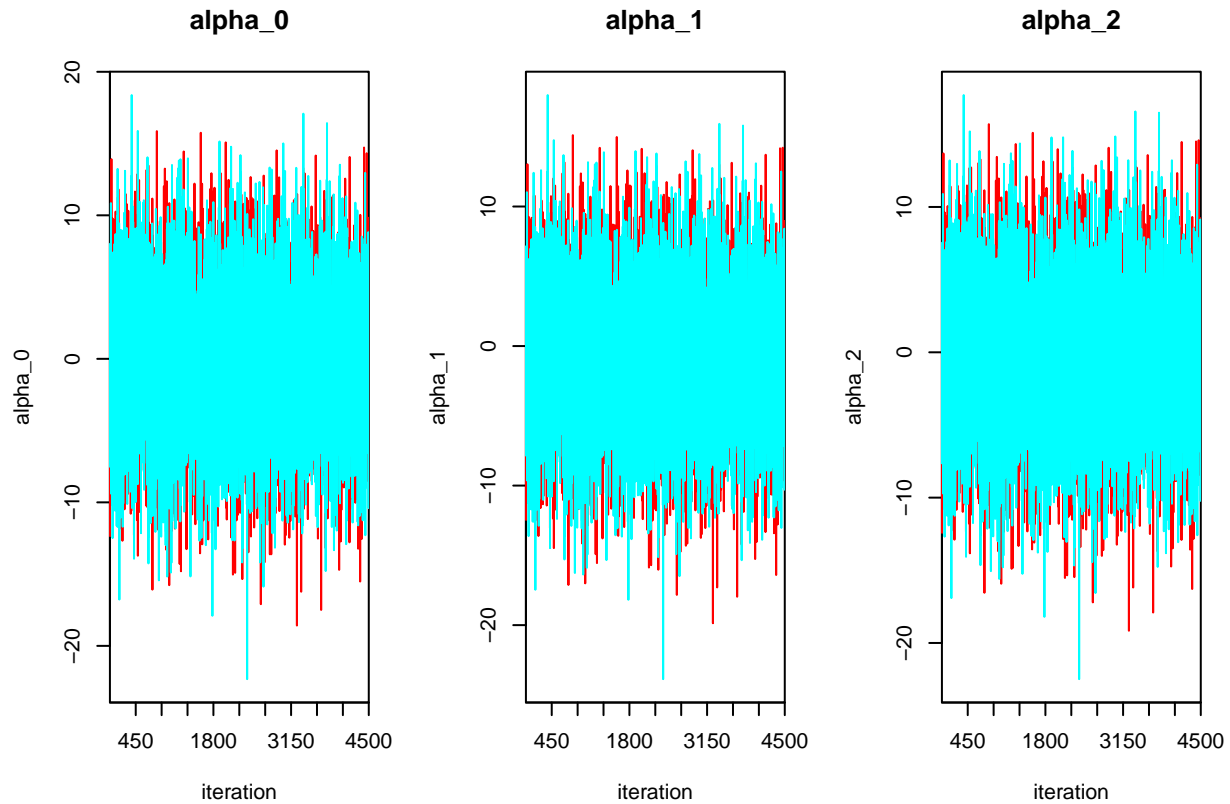
M3

```

## Inference for Bugs model at "C:/Users/basil/AppData/Local/Temp/RtmpaCcf2h/model79e84a471826.txt", fi
## 2 chains, each with 9000 iterations (first 4500 discarded)
## n.sims = 9000 iterations saved
##      mu.vect sd.vect   2.5%   25%   50%   75%  97.5%  Rhat n.eff
## alpha_0  -0.434   4.993 -10.078 -3.885 -0.536  2.932  9.318 1.001  9000
## alpha_1  -1.092   4.996 -10.616 -4.498 -1.194  2.285  8.669 1.001  9000
## alpha_2  -0.632   4.995 -10.255 -4.072 -0.734  2.720  9.195 1.001  9000
## mu.0      -2.115   4.992 -11.907 -5.475 -2.025  1.319  7.469 1.001  9000
## deviance 103.133   7.234 100.282 101.263 102.398 104.080 109.047 1.045  9000
##
## For each parameter, n.eff is a crude measure of effective sample size,

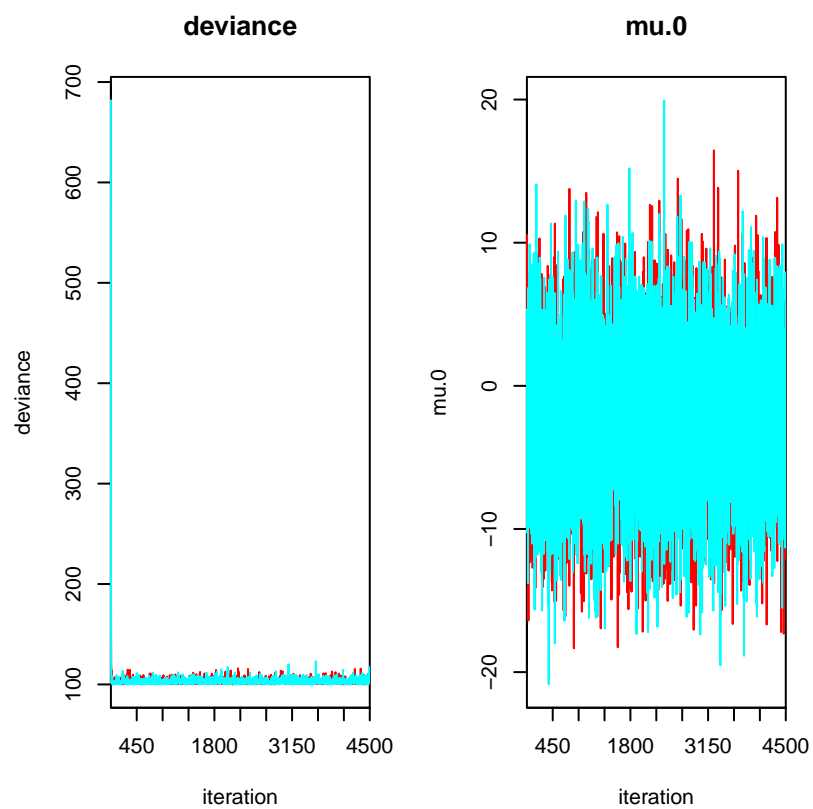
```

```
## and Rhat is the potential scale reduction factor (at convergence, Rhat=1).
##
## DIC info (using the rule,  $pD = \text{var}(\text{deviance})/2$ )
##  $pD = 26.2$  and  $DIC = 129.3$ 
## DIC is an estimate of expected predictive error (lower deviance is better).
traceplot(M3, mfrow=c(1, 3), ask=FALSE)
```



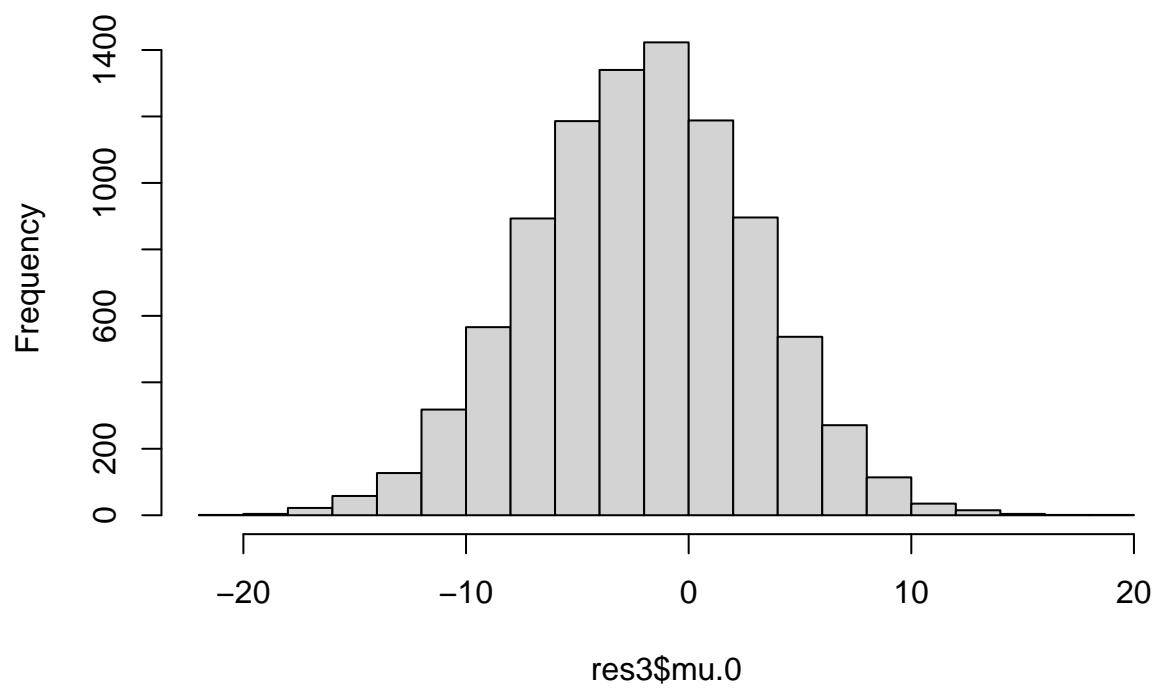
```
par(mfrow = c(1, 1))
```





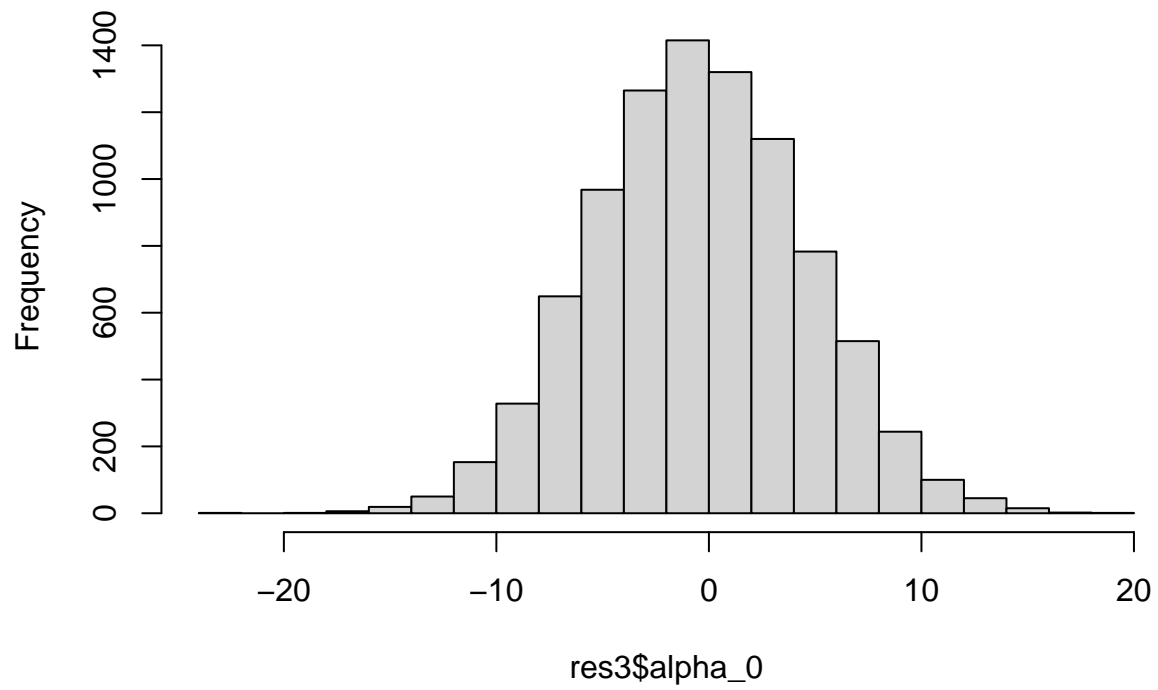
```
res3 <- M3$BUGSoutput$sims.matrix %>%  
  as.data.frame()  
hist(res3$mu.0)
```

**Histogram of res3\$mu.0**

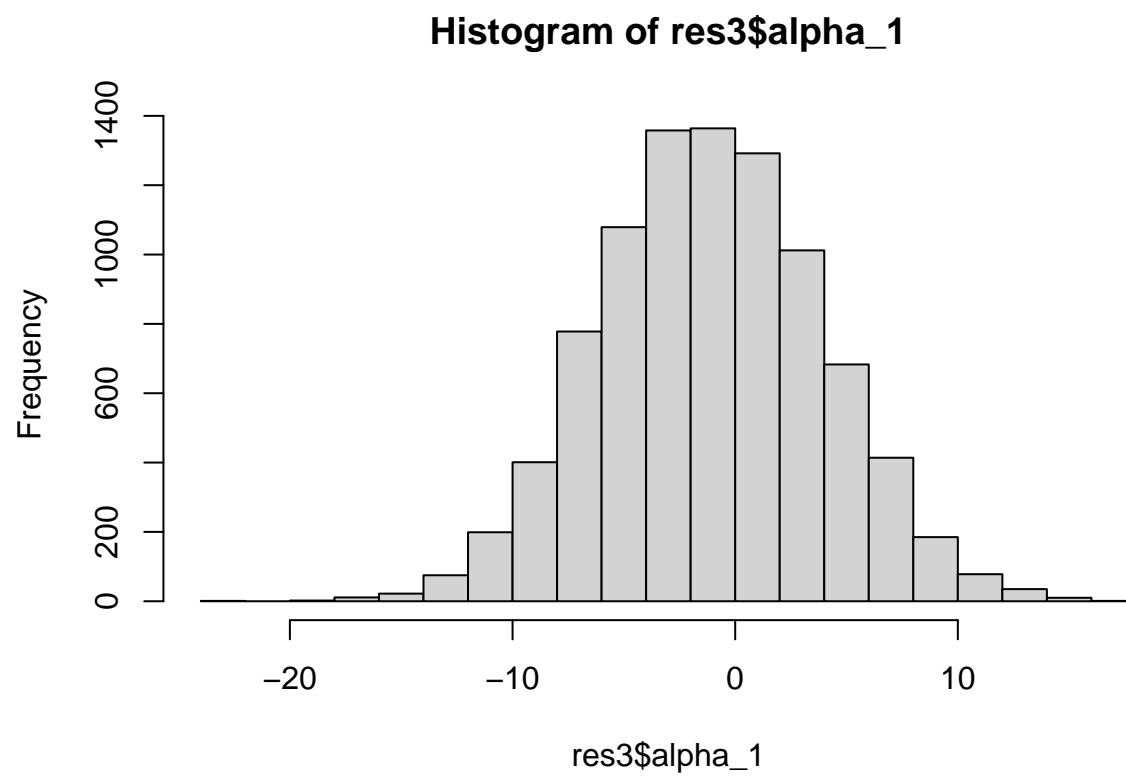


```
hist(res3$alpha_0)
```

**Histogram of res3\$alpha\_0**

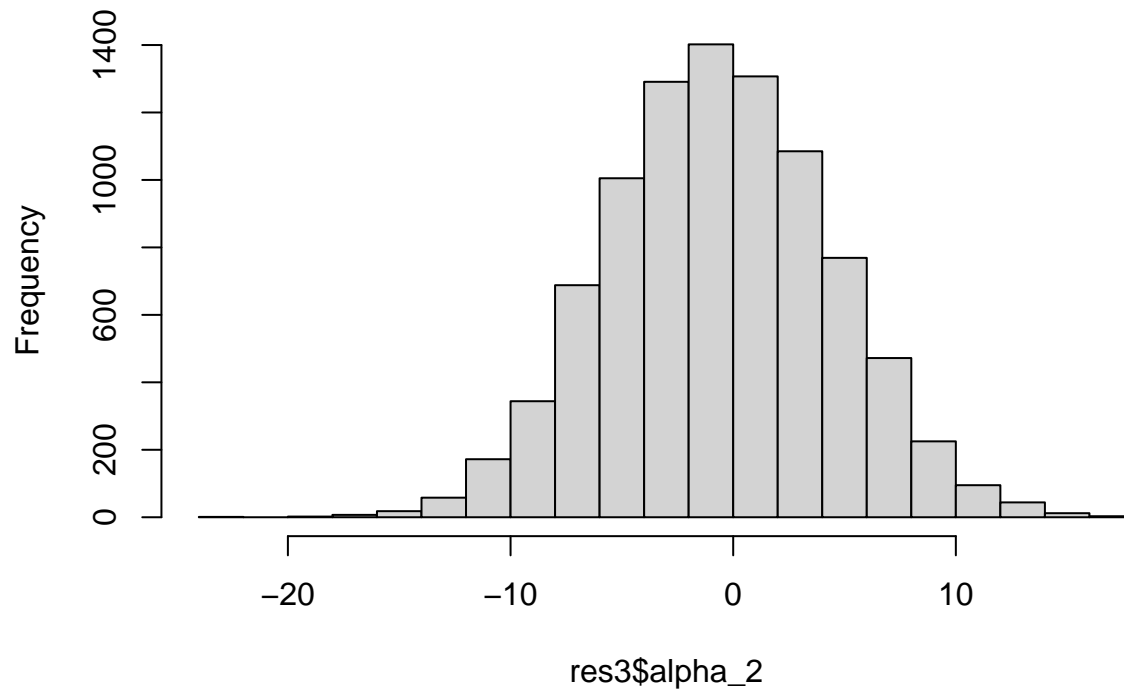


```
hist(res3$alpha_1)
```



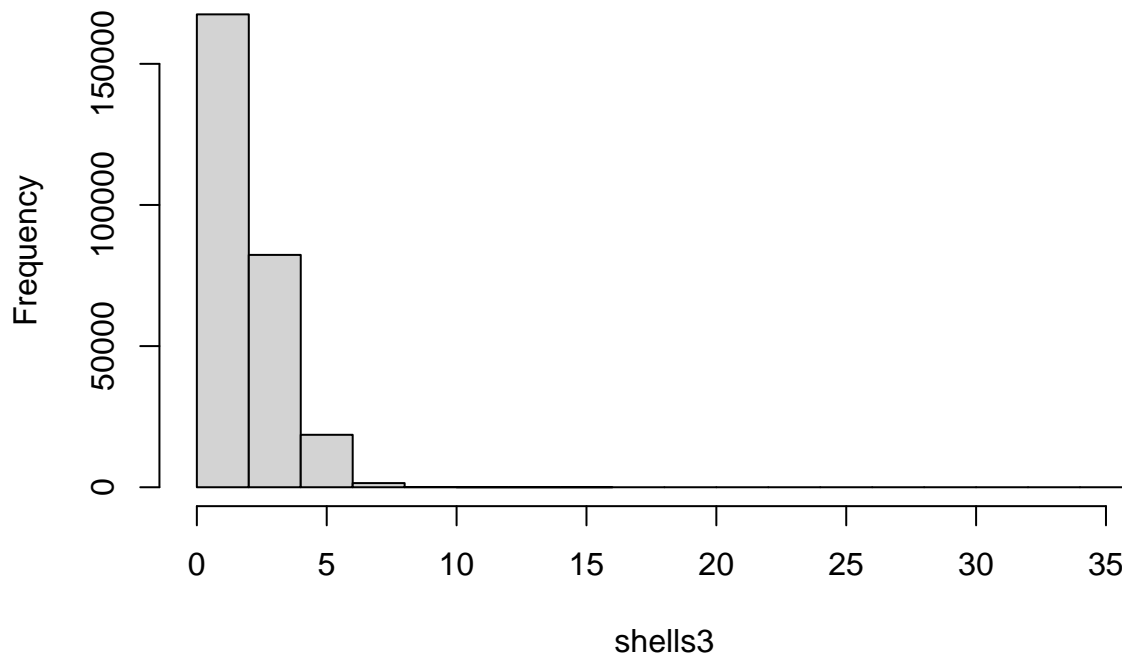
```
hist(res3$alpha_2)
```

## Histogram of res3\$alpha\_2



```
# Rétrotransformation
shells3 <- matrix(NA, ncol=nrow(gopher), nrow=nrow(res3))
for (i in 1:nrow(gopher)){
  shells3[,i] <- gopher$Area[i] * exp(res3$mu.0 + res3$alpha_0 * gopher$Cov_2004[i] + res3$alpha_1 * gopher$Cov_2004[i]^2)
}
hist(shells3)
```

## Histogram of shells3



```
mean(shells3)
```

```
## [1] 1.817698
```

```
# Récupérons le DIC
```

```
DIC3 <- M3$BUGSoutput$DIC
```

Un modèle avec l'année et la prévalence (continue) M4 :

$$\log(\lambda_i) = \log(aire_i) + \mu_0 + \alpha_0 * cov_{2004} + \alpha_1 * cov_{2005} + \alpha_2 * cov_{2006} + \beta * prev_i$$

```
# The year model function
```

```
year_prevalence_model <- function(){
```

```
# This model takes into account the variable year with it's 3 modalities.
```

```
# Likelihood
```

```
for (i in 1:N){ # Loop over observations
```

```
  S[i] ~ dpois(lambda[i])
```

```
  log(lambda[i]) <- mu.0 + alpha_0 * Cov_2004[i] + alpha_1 * Cov_2005[i] + alpha_2 * Cov_2006[i] + b.
```

```
}
```

```
# Priors
```

```
mu.0 ~ dnorm(0, 1/100)
```

```
b.prev ~ dnorm(0, 1/100)
```

```
alpha_0 ~ dnorm(0, 1/100)
```

```
alpha_1 ~ dnorm(0, 1/100)
```

```
alpha_2 ~ dnorm(0, 1/100)
```

```
}
```

```
# Now, we can make list of the data to use in the jags function
```

```

datax <- list(
  N = gopher$year %>%
    length(),
  S = gopher$shells,      # Number of shells
  A = gopher$Area,        # Area offset
  prev=gopher$standprev,  # Standardized prev
  Cov_2004 = gopher$Cov_2004, # Effect of the year 2004
  Cov_2005 = gopher$Cov_2005, # Effect of the year 2005
  Cov_2006 = gopher$Cov_2006 # Effect of the year 2006
)

# Parameters to estimate
params <- c("mu.0", "alpha_0", "alpha_1", "alpha_2", "b.prev")

# Initialising the chains
init1 <- list(mu.0 = -0.5, alpha_0=-0.5, alpha_1=-0.5, alpha_2=-0.5, b.prev =-0.5)
init2 <- list(mu.0 = 0.5, alpha_0=0.5, alpha_1=0.5, alpha_2=0.5, b.prev=0.5)
init <- list(init1, init2)

# Define the iteration parameters
nb.iterations <- 9000
nb.burnin <- 4500
# Run the model using jags
M4 <- jags(
  data = datax,
  inits = init,
  parameters.to.save = params,
  model.file = year_prevalence_model,
  n.chains = 2,
  n.iter = nb.iterations,
  n.burnin = nb.burnin,
  n.thin = 1
)

```

```

## Compiling model graph
##   Resolving undeclared variables
##   Allocating nodes
## Graph information:
##   Observed stochastic nodes: 30
##   Unobserved stochastic nodes: 5
##   Total graph size: 289
##
## Initializing model

```

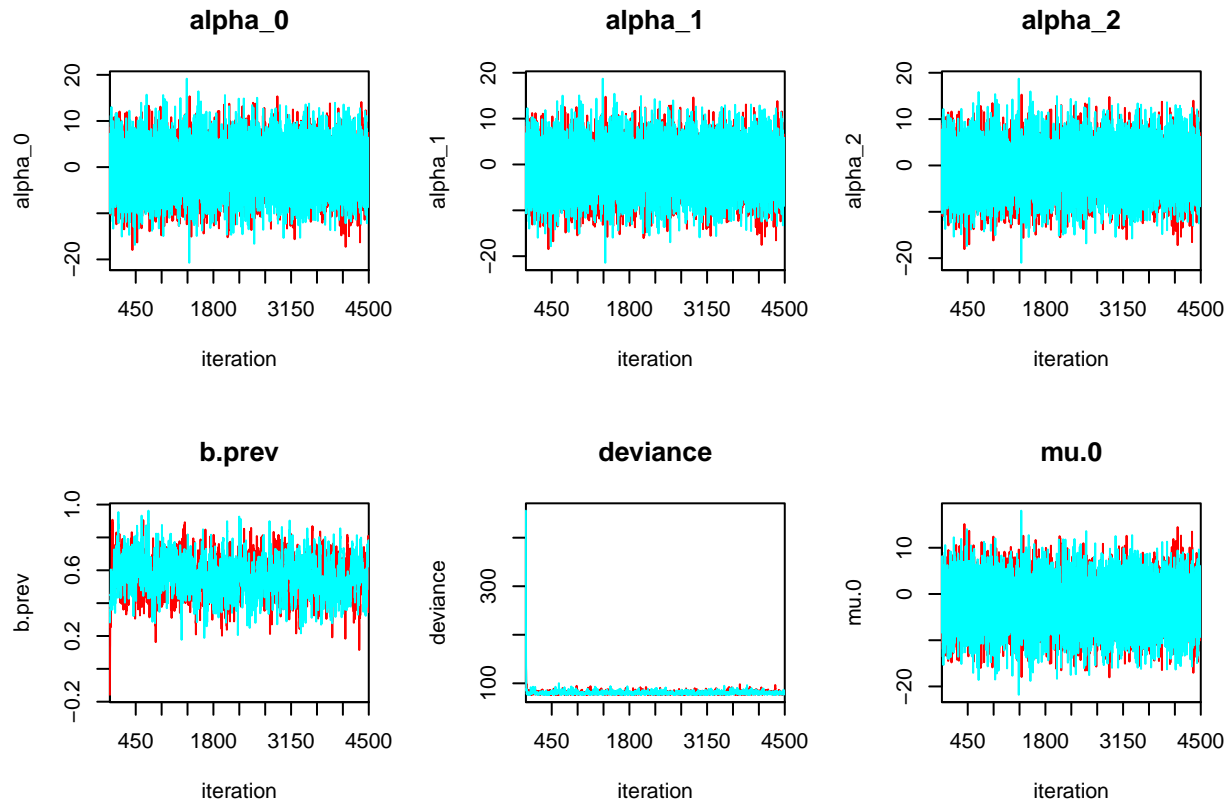
M4

```

## Inference for Bugs model at "C:/Users/basil/AppData/Local/Temp/RtmpaCcf2h/model79e84dd14bce.txt", fi
## 2 chains, each with 9000 iterations (first 4500 discarded)
## n.sims = 9000 iterations saved
##      mu.vect sd.vect   2.5%   25%   50%   75%  97.5%  Rhat n.eff
## alpha_0  -0.339   5.030 -10.303 -3.751 -0.375  2.950  9.710 1.001  6000
## alpha_1  -1.015   5.028 -10.949 -4.405 -1.039  2.291  9.089 1.001  7500
## alpha_2  -0.756   5.031 -10.629 -4.130 -0.798  2.529  9.249 1.001  4200
## b.prev    0.561   0.114   0.339  0.484  0.563  0.639  0.778 1.001  9000
## mu.0      -2.355   5.026 -12.398 -5.663 -2.320  1.051  7.577 1.001  4700

```

```
## deviance 80.390 5.506 76.811 78.232 79.616 81.648 87.538 1.014 9000
##
## For each parameter, n.eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor (at convergence, Rhat=1).
##
## DIC info (using the rule, pD = var(deviance)/2)
## pD = 15.2 and DIC = 95.5
## DIC is an estimate of expected predictive error (lower deviance is better).
traceplot(M4, mfrow=c(2, 3), ask=FALSE)
```

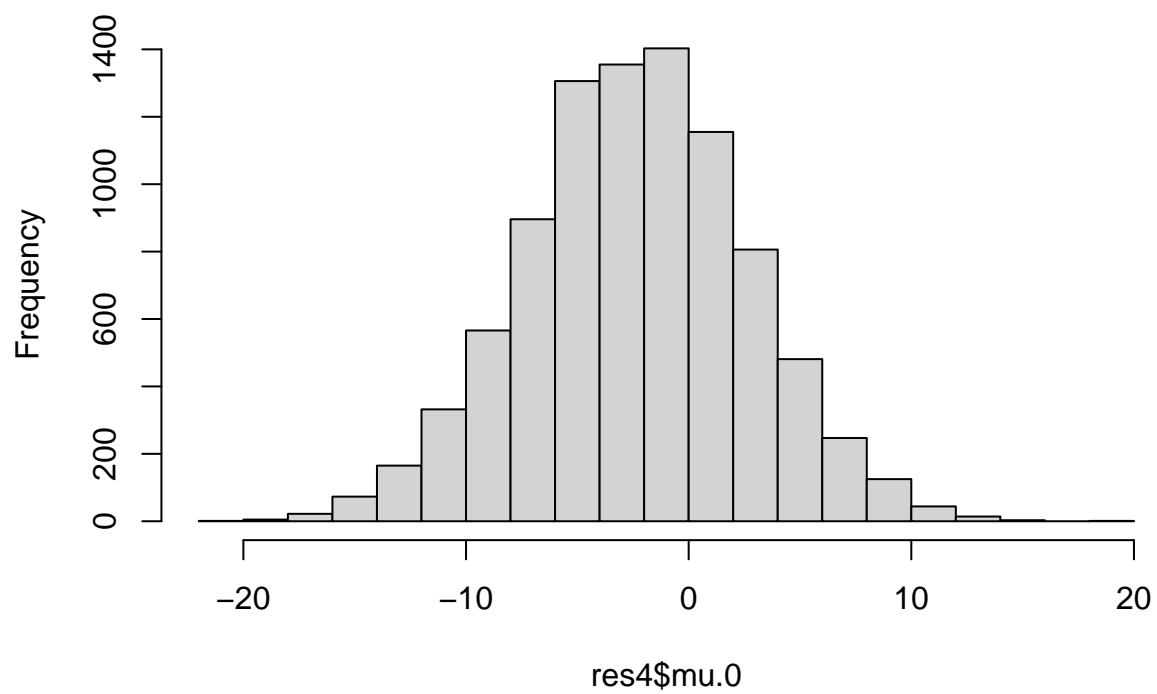


```
par(mfrow = c(1, 1))

res4 <- M4$BUGSoutput$sims.matrix %>%
  as.data.frame()
hist(res4$mu.0)
```

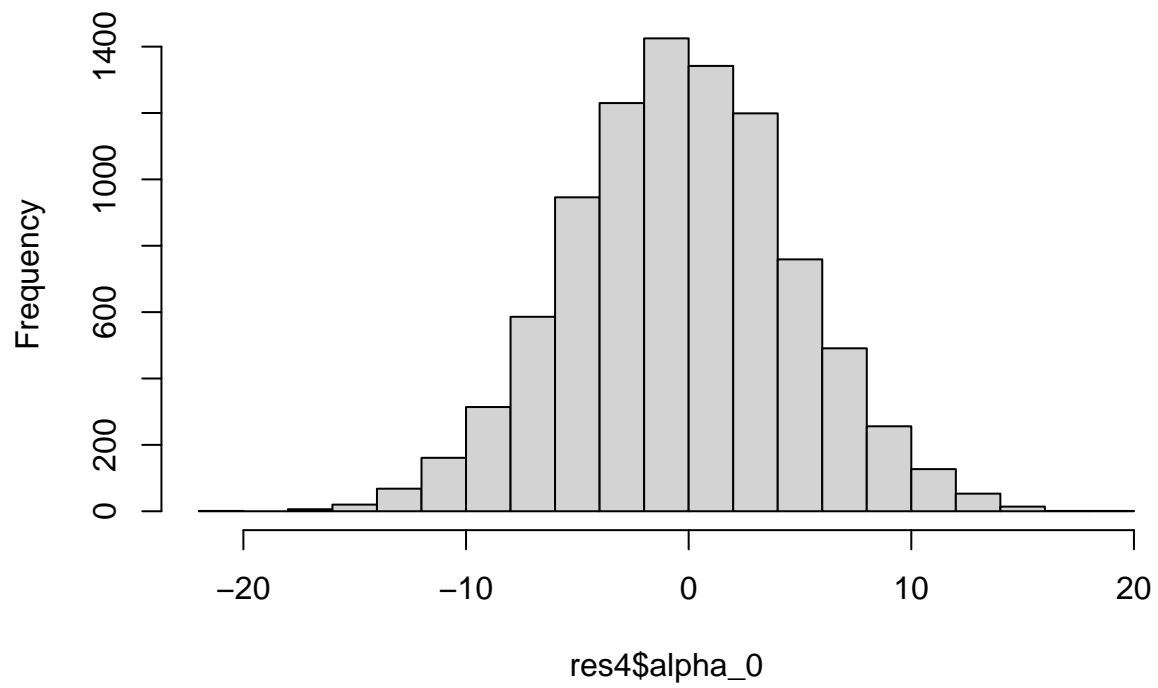


**Histogram of res4\$mu.0**

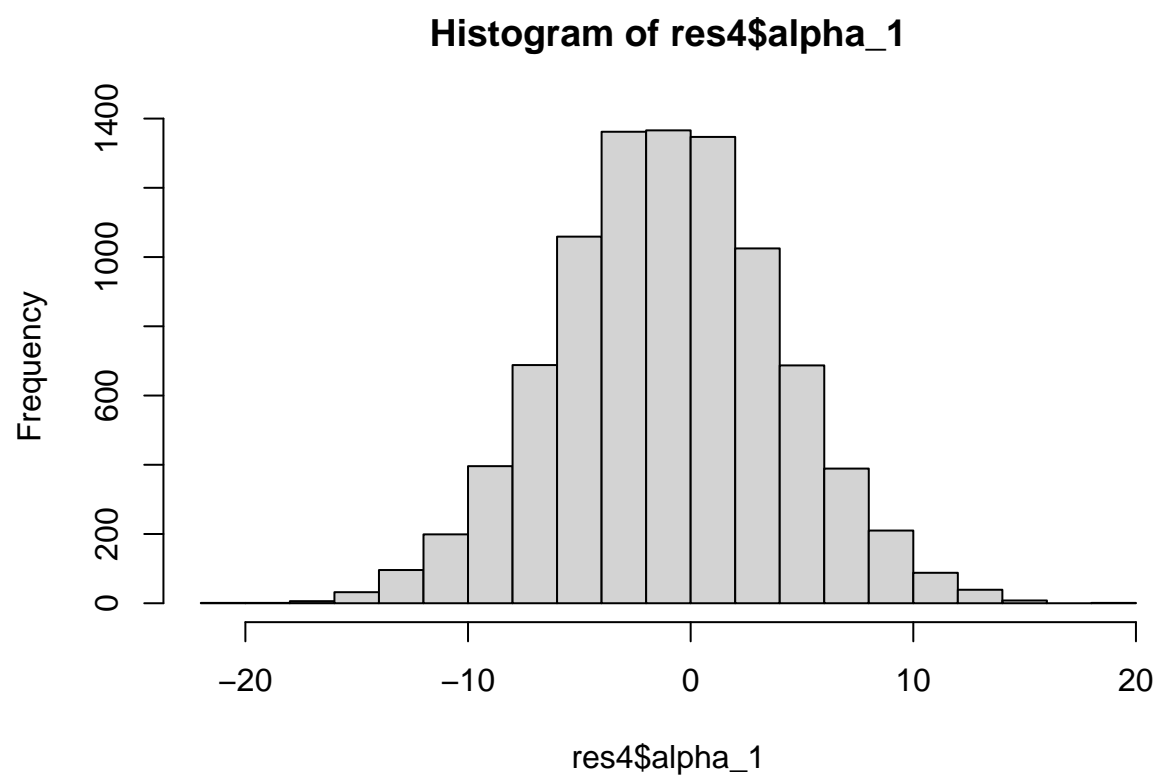


```
hist(res4$alpha_0)
```

**Histogram of res4\$alpha\_0**

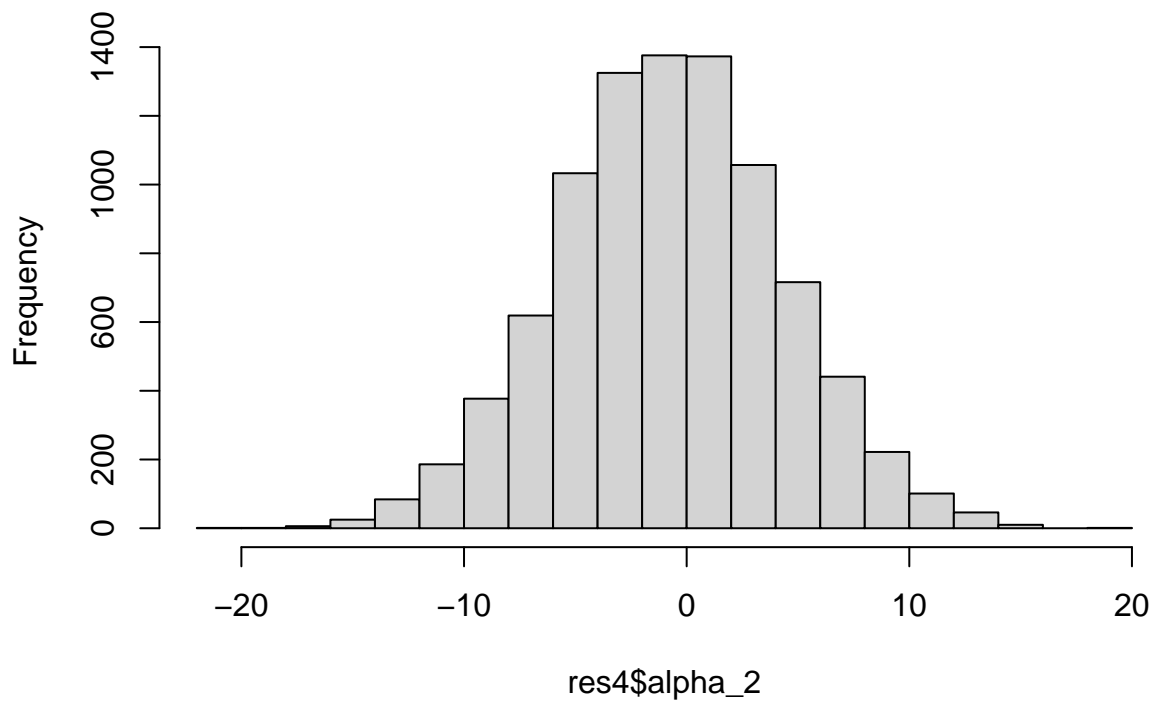


```
hist(res4$alpha_1)
```



```
hist(res4$alpha_2)
```

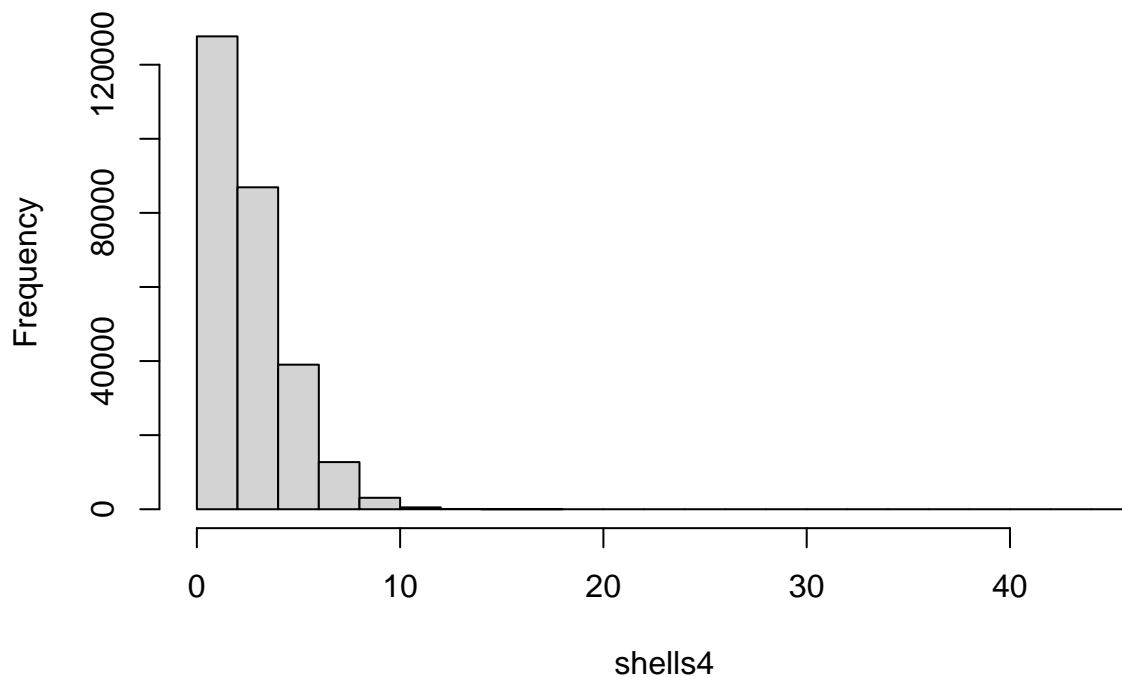
**Histogram of res4\$alpha\_2**



```
# Rétrotransformation
shells4 <- matrix(NA, ncol=nrow(gopher), nrow=nrow(res4))
names(res4)

## [1] "alpha_0" "alpha_1" "alpha_2" "b.prev" "deviance" "mu.0"
for (i in 1:nrow(gopher)){
  shells4[,i] <- gopher$Area[i] * exp(res4$mu.0 + res4$alpha_0 * gopher$Cov_2004[i] + res4$alpha_1 * gopher$Cov_2004[i] + res4$alpha_2 * gopher$Cov_2004[i])
}
hist(shells4)
```

## Histogram of shells4



```
mean(shells4)
```

```
## [1] 2.557749
```

```
# Récupérons le DIC
```

```
DIC4 <- M4$BUGSoutput$DIC
```

Modèle avec année et site M5 :

$$\log(\lambda_i) = \log(\text{aire}_i) + \mu_0 + \alpha_0 * \text{cov}_{2004} + \alpha_1 * \text{cov}_{2005} + \alpha_2 * \text{cov}_{2006} + \gamma_j \text{ avec } j = 1, \dots, 10$$

```
year_site_model <- function(){
  # This model takes into account a random effect for the site
  # Likelihood
  for(i in 1:N){
    S[i] ~ dpois(lambda[i])
    log(lambda[i]) <- mu.0 + gamma[site[i]] + alpha_0* Cov_2004[i] + alpha_1 * Cov_2005[i] + alpha_2 * C
  }
  for (j in 1:nb.sites){
    gamma[j] ~ dnorm(0, tau.s)
  }
  # Priors
  mu.0 ~ dnorm(0, 0.001)
  sd.s ~ dunif(0, 100)
  tau.s <- 1 / (sd.s * sd.s)
  alpha_0 ~ dnorm(0, 1/100)
  alpha_1 ~ dnorm(0, 1/100)
  alpha_2 ~ dnorm(0, 1/100)
}
```

```

}

# Make the data to use in jags
datax <- list(
  N = gopher$year %>%
    length(),
  S = gopher$shells,
  A = gopher$Area,
  site = gopher$Site %>%
    as.numeric(),
  nb.sites = gopher$Site %>%
    unique() %>%
    length(),
  Cov_2004 = ifelse(gopher$year == 2004, 1, 0),
  Cov_2005 = ifelse(gopher$year == 2005, 1, 0),
  Cov_2006 = ifelse(gopher$year == 2006, 1, 0)
)

# Make a list of parameters to save
params = c("mu.0", "sd.s", "alpha_0", "alpha_1", "alpha_2")

# Initial conditions
init1 <- list(
  "mu.0" = 0.5,
  "alpha_0" = 0.5,
  "alpha_1" = 0.5,
  "alpha_2" = 0.5,
  "sd.s" = 0.5
)
init2 <- list(
  "mu.0" = -0.5,
  "alpha_0" = -0.5,
  "alpha_1" = -0.5,
  "alpha_2" = -0.5,
  "sd.s" = 1.5
)
init <- list(init1, init2)

# Iteration parameters
nb.iterations <- 9000
nb.burnin <- 4500

# Run the model
M5 <- jags(
  data = datax,
  parameters.to.save = params,
  inits = init,
  model.file = year_site_model,
  n.chains = 2,
  n.iter = nb.iterations,
  n.burnin = nb.burnin,
  n.thin = 1
)

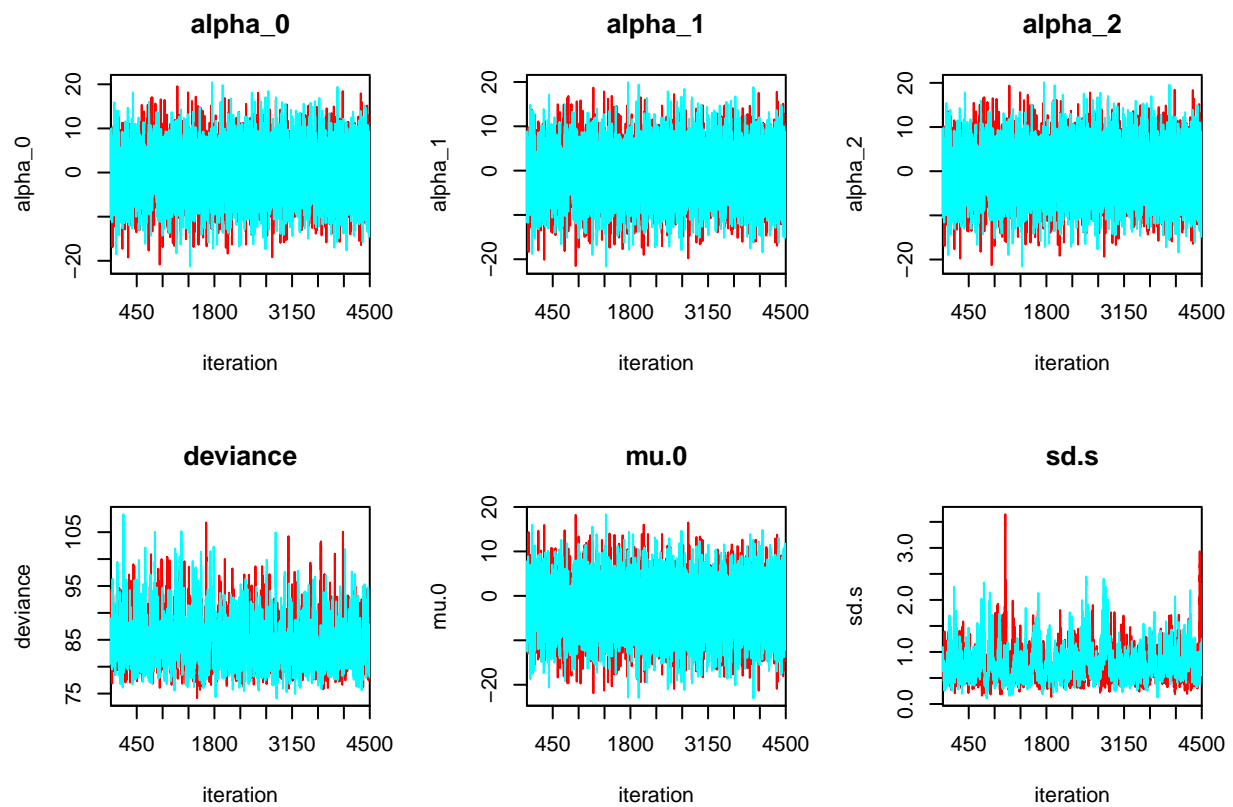
```

```

## Compiling model graph
##   Resolving undeclared variables
##   Allocating nodes
## Graph information:
##   Observed stochastic nodes: 30
##   Unobserved stochastic nodes: 15
##   Total graph size: 280
##
## Initializing model
# Regardons le modèle et les traces
M5

## Inference for Bugs model at "C:/Users/basil/AppData/Local/Temp/RtmpaCcf2h/model79e8672b7875.txt", fi
## 2 chains, each with 9000 iterations (first 4500 discarded)
## n.sims = 9000 iterations saved
##      mu.vect sd.vect   2.5%   25%   50%   75%  97.5%  Rhat n.eff
## alpha_0    0.104   5.718 -11.146 -3.743  0.136  3.934 11.193 1.001  9000
## alpha_1   -0.520   5.725 -11.728 -4.396 -0.496  3.349 10.596 1.001  9000
## alpha_2   -0.079   5.719 -11.283 -3.924 -0.058  3.789 11.027 1.001  9000
## mu.0       -2.909   5.727 -14.020 -6.728 -2.936  0.959  8.321 1.001  9000
## sd.s        0.742   0.323  0.301  0.520  0.684  0.896  1.541 1.010   160
## deviance   84.467   4.393  77.647 81.259 83.861 87.002 94.617 1.004   420
##
## For each parameter, n.eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor (at convergence, Rhat=1).
##
## DIC info (using the rule, pD = var(deviance)/2)
## pD = 9.6 and DIC = 94.1
## DIC is an estimate of expected predictive error (lower deviance is better).
traceplot(M5, mfrow=c(2, 3), ask=FALSE)

```

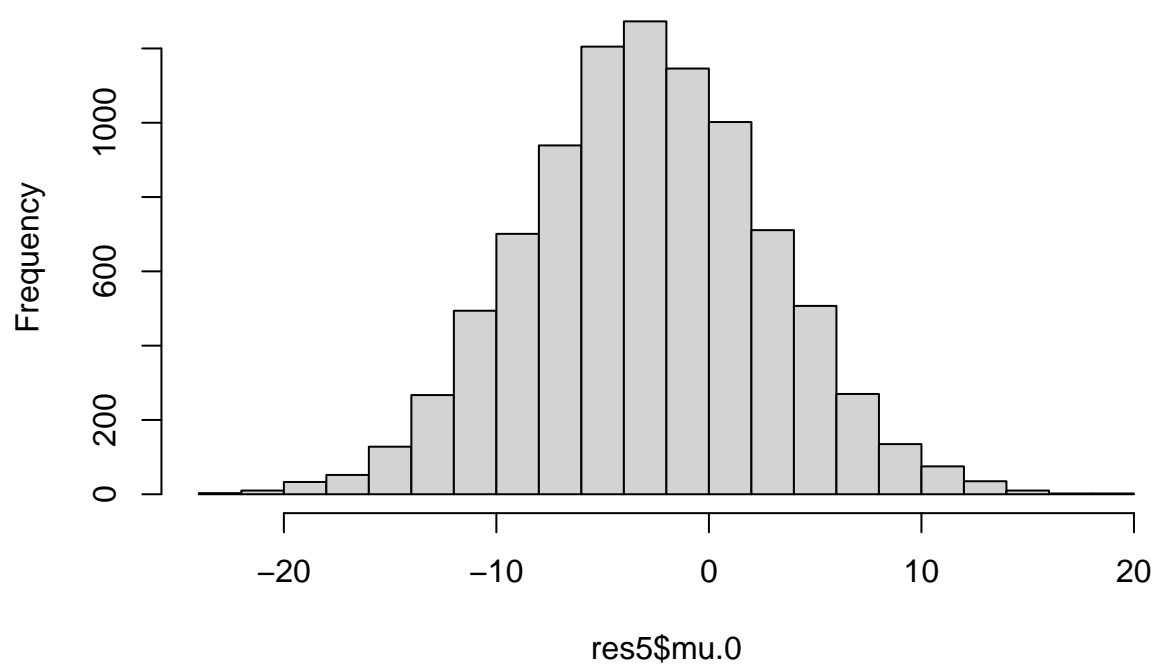


```
par(mfrow=c(1, 1))

# On récupère les paramètres et on regarde leur distribution
res5 <- M5$BUGSoutput$sims.matrix %>%
  as.data.frame()
hist(res5$mu.0)
```

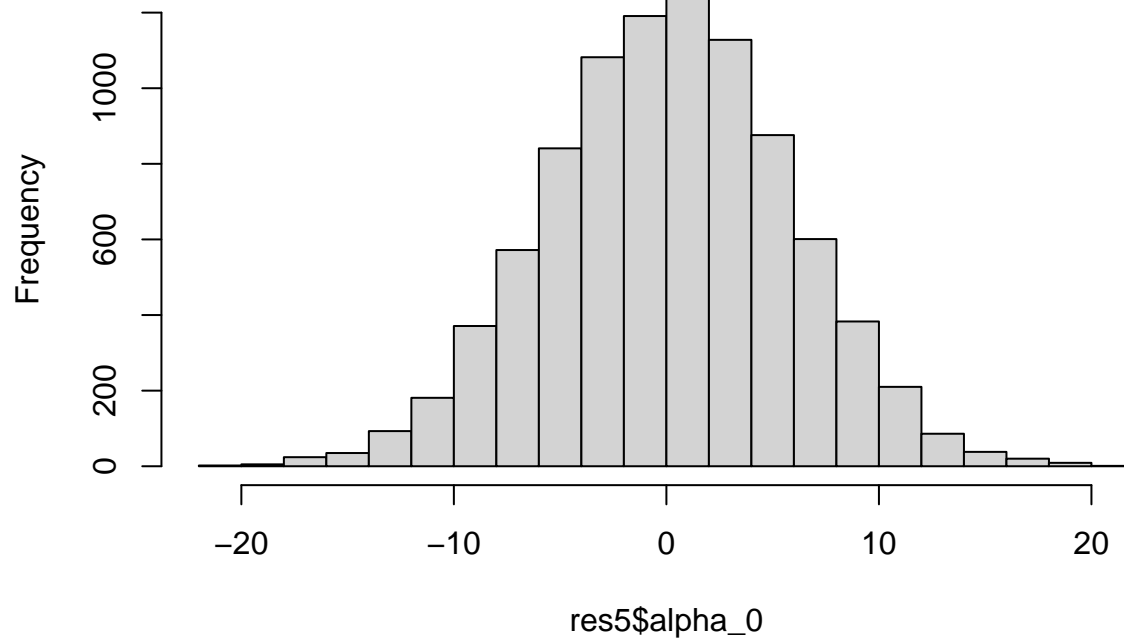


**Histogram of res5\$mu.0**



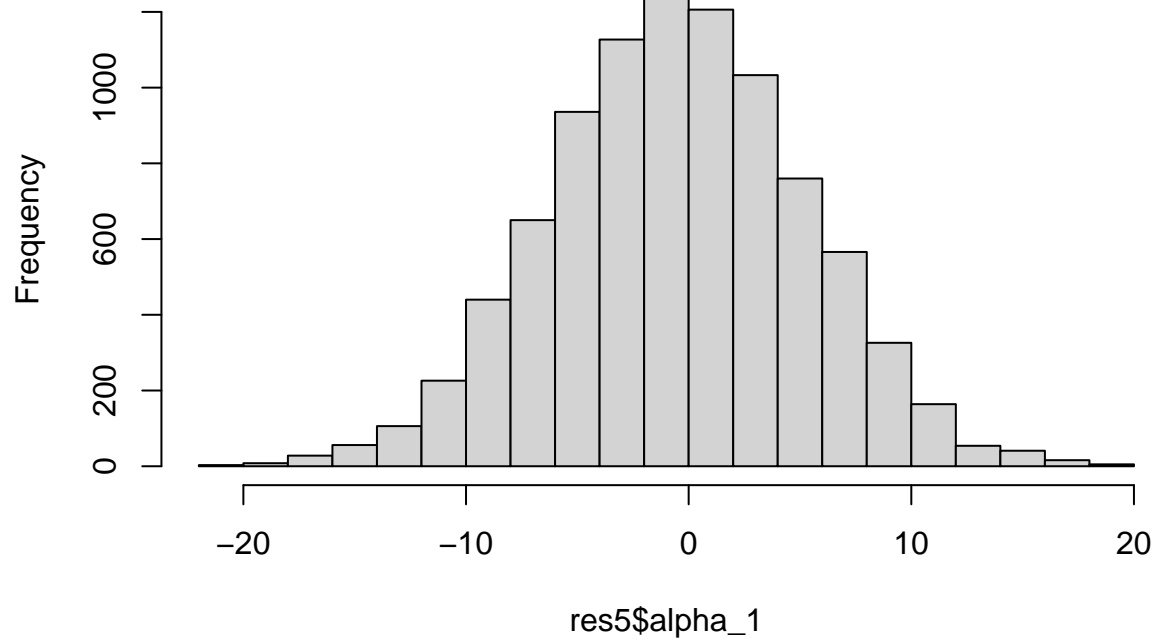
```
hist(res5$alpha_0)
```

**Histogram of res5\$alpha\_0**



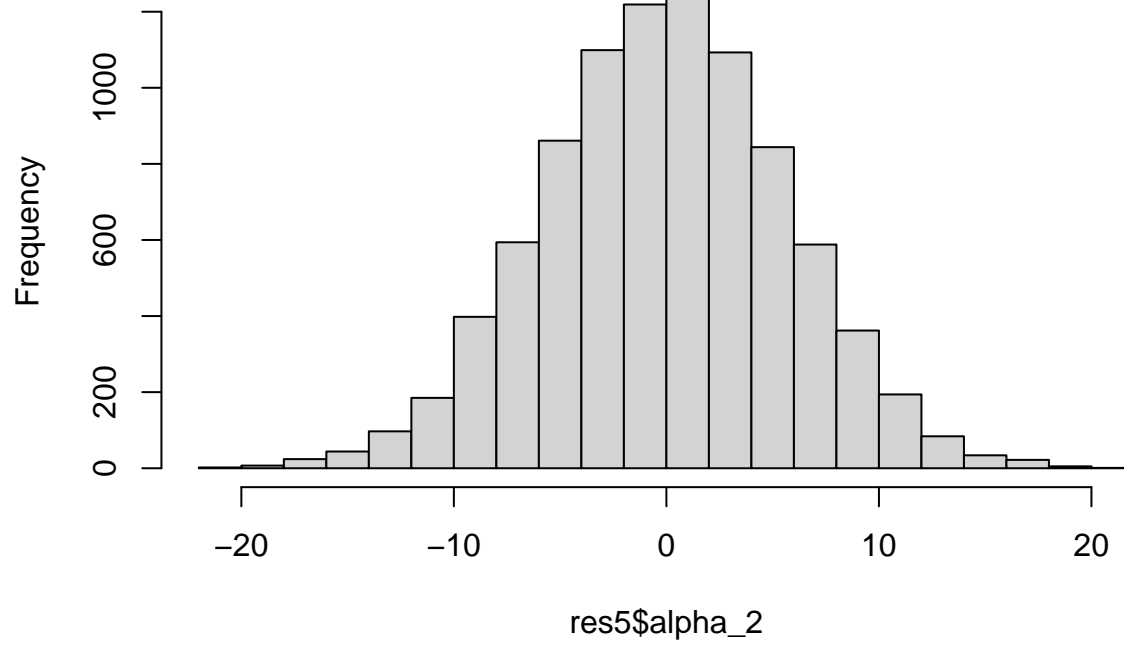
```
hist(res5$alpha_1)
```

**Histogram of res5\$alpha\_1**



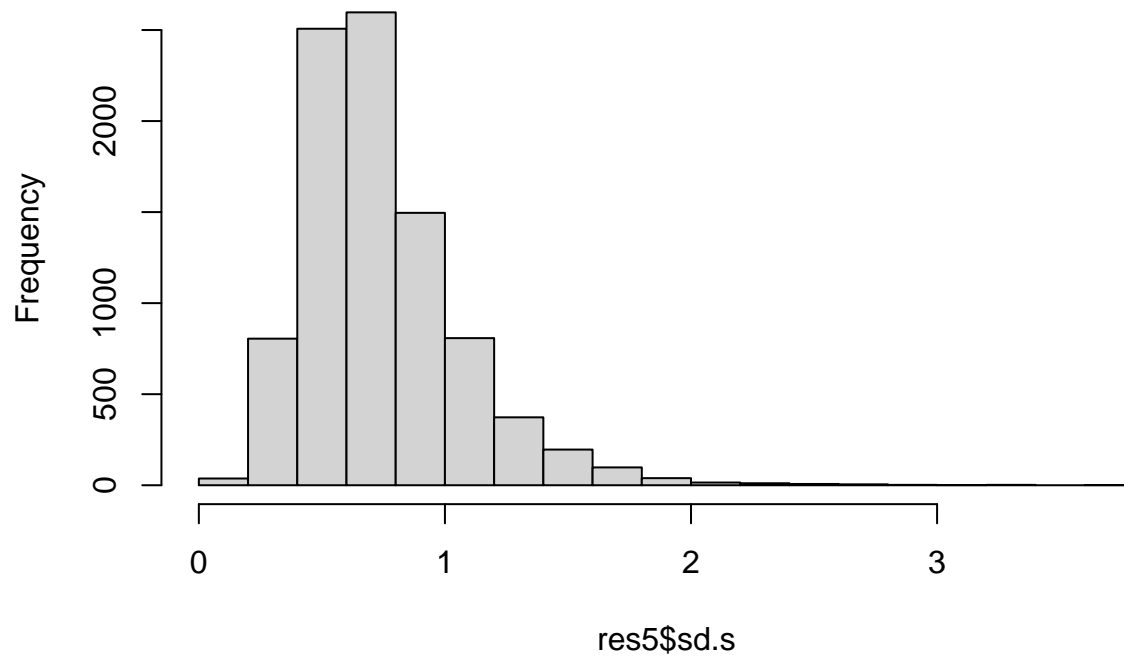
```
hist(res5$alpha_2)
```

**Histogram of res5\$alpha\_2**



```
hist(res5$sd.s)
```

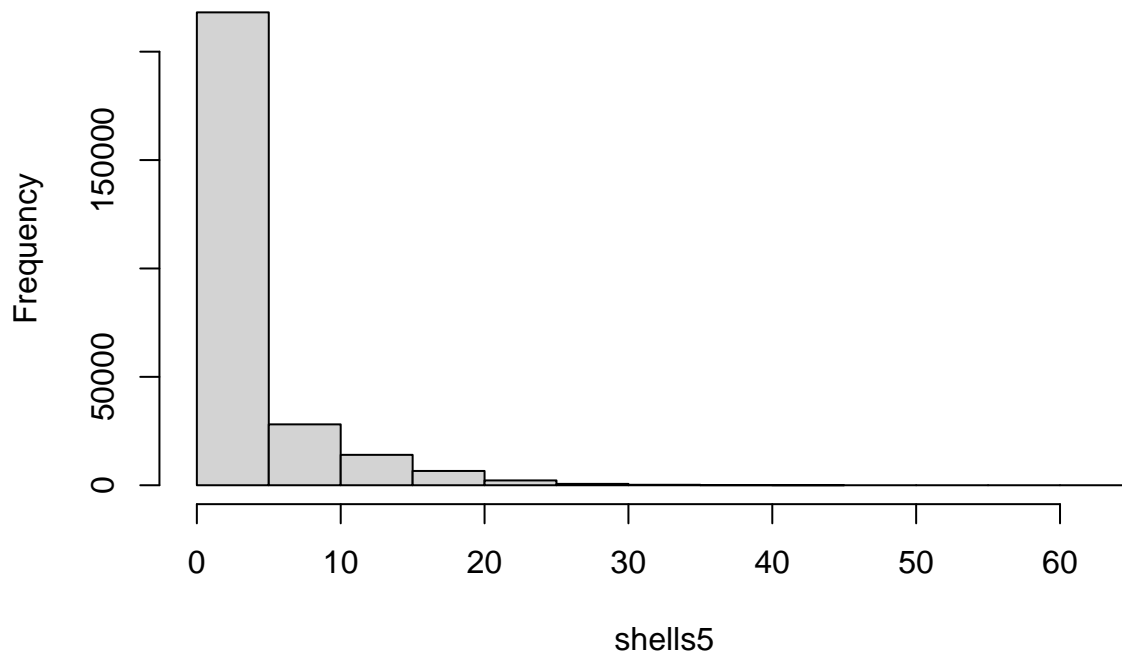
## Histogram of res5\$sd.s



```
# Rétrotransformation
```

```
shells5 <- matrix(NA, ncol=nrow(gopher), nrow=nrow(res5))  
for (i in 1:nrow(gopher)){  
  shells5[,i] <- gopher$Area[i] * exp(res5$mu.0 + res5$alpha_0 * gopher$Cov_2004[i] + res5$alpha_1 * gopher$Cov_2004[i])  
}  
hist(shells5)
```

## Histogram of shells5



```
mean(shells5)
```

```
## [1] 2.997636
```

```
# Récupérons le DIC
```

```
DIC5 <- M5$BUGSoutput$DIC
```

Partir directement avec les effets aléatoires (prendre en compte la structuration des données)

## 4 Inférence et interprétation des résultats

Sur la base du meilleur modèle, donnez les estimations des paramètres ainsi qu'une mesure de l'incertitude associée. Interprétez vos résultats.

## 5 Discussion

Comparez vos résultats à ceux du papier. Sont-ils semblables ou différents? Pourquoi selon vous? Si cela vous semble pertinent, proposez des pistes d'amélioration de l'analyse.

```
random_model <- function(){  
  # This model takes into account a random effect for the site  
  # Likelihood  
  for(i in 1:N){  
    S[i] ~ dpois(lambda[i])  
    log(lambda[i]) <- mu.0 + gamma[site[i]] + b.prev * prev[i] + alpha_0* Cov_2004[i] + alpha_1 * Cov_2005[i]  
  }  
  for (j in 1:nb.sites){
```

```

    gamma[j] ~ dnorm(0, tau.s)
  }
  # Priors
  mu.0 ~ dnorm(0, 0.001)
  sd.s ~ dunif(0, 100)
  tau.s <- 1 / (sd.s * sd.s)
  b.prev ~ dnorm(0, 1/100)
  alpha_0 ~ dnorm(0, 1/100)
  alpha_1 ~ dnorm(0, 1/100)
  alpha_2 ~ dnorm(0, 1/100)
}

# Make the data to use in jags
datax <- list(
  N = gopher$year %>%
    length(),
  S = gopher$shells,
  prev = gopher$standprev,
  A = gopher$Area,
  site = gopher$Site %>%
    as.numeric(),
  nb.sites = gopher$Site %>%
    unique() %>%
    length(),
  Cov_2004 = ifelse(gopher$year == 2004, 1, 0),
  Cov_2005 = ifelse(gopher$year == 2005, 1, 0),
  Cov_2006 = ifelse(gopher$year == 2006, 1, 0)
)

# Make a list of parameters to save
params = c("mu.0", "b.prev", "sd.s", "alpha_0", "alpha_1", "alpha_2")

# Initial conditions
init1 <- list(
  "mu.0" = 0.5,
  "b.prev" = 0.5,
  "alpha_0" = 0.5,
  "alpha_1" = 0.5,
  "alpha_2" = 0.5,
  "sd.s" = 0.5
)
init2 <- list(
  "mu.0" = - 0.5,
  "b.prev" = - 0.5,
  "alpha_0" = -0.5,
  "alpha_1" = -0.5,
  "alpha_2" = -0.5,
  "sd.s" = 1.5
)
init <- list(init1, init2)

# Iteration parameters
nb.iterations <- 9000

```

```
nb.burnin <- 4500
```

```
# Run the model
```

```
M6 <- jags(  
  data = datax,  
  parameters.to.save = params,  
  inits = init,  
  model.file = random_model,  
  n.chains = 2,  
  n.iter = nb.iterations,  
  n.burnin = nb.burnin,  
  n.thin = 1  
)
```

```
## Compiling model graph  
##   Resolving undeclared variables  
##   Allocating nodes  
## Graph information:  
##   Observed stochastic nodes: 30  
##   Unobserved stochastic nodes: 16  
##   Total graph size: 334  
##  
## Initializing model
```

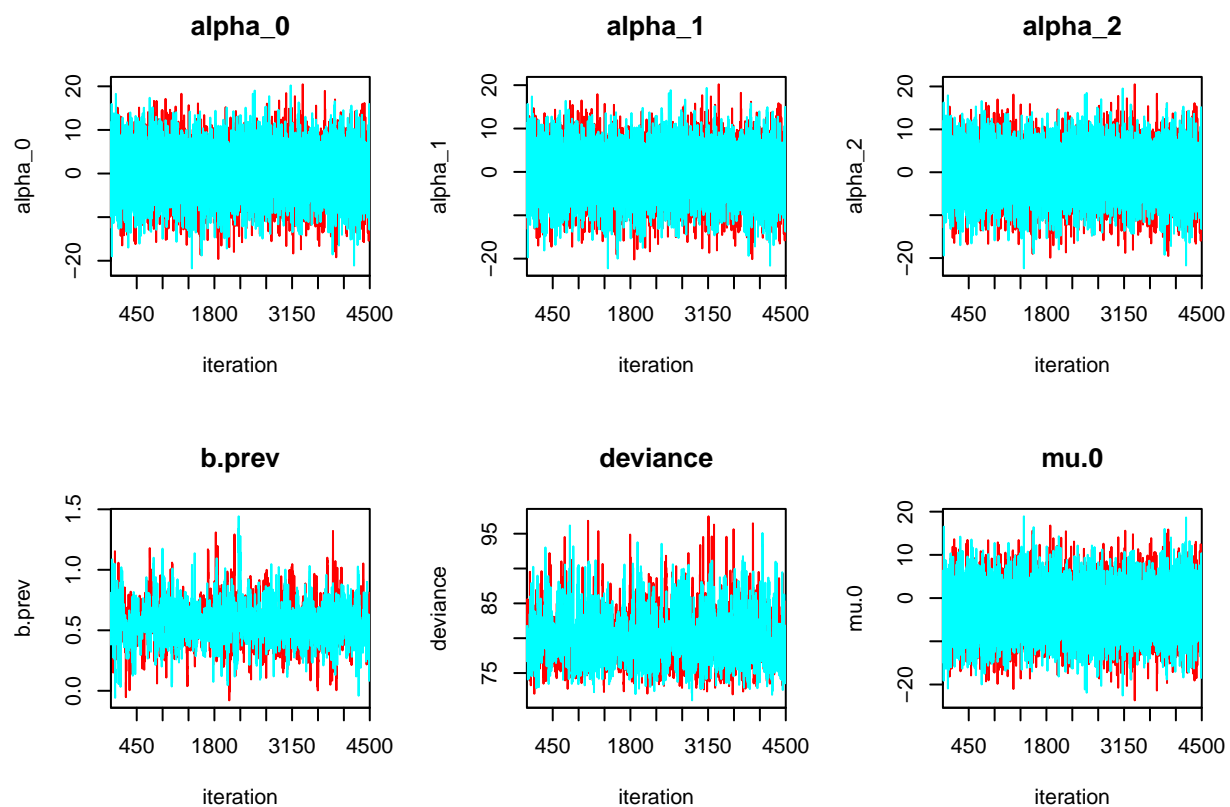
```
# Regardons le modèle et les traces
```

```
M6
```

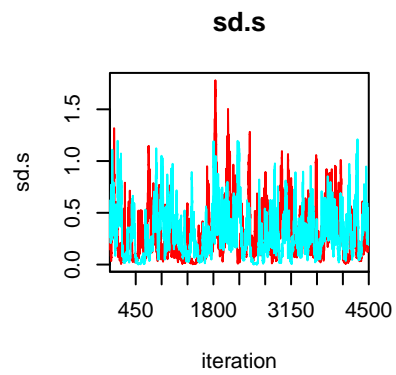
```
## Inference for Bugs model at "C:/Users/basil/AppData/Local/Temp/RtmpaCcf2h/model79e8458296.txt", fit v  
## 2 chains, each with 9000 iterations (first 4500 discarded)  
## n.sims = 9000 iterations saved  
##      mu.vect sd.vect   2.5%   25%   50%   75%  97.5%  Rhat n.eff  
## alpha_0    0.235   5.721 -11.219 -3.638  0.171  4.214 11.241 1.001  9000  
## alpha_1   -0.444   5.728 -11.993 -4.328 -0.510  3.541 10.558 1.001  9000  
## alpha_2   -0.193   5.722 -11.500 -4.047 -0.258  3.804 10.767 1.001  9000  
## b.prev     0.562   0.163  0.248  0.460  0.557  0.658  0.909 1.001  9000  
## mu.0       -2.947   5.724 -13.942 -6.941 -2.875  0.909  8.527 1.001  9000  
## sd.s        0.300   0.235  0.013  0.119  0.249  0.426  0.895 1.003  7400  
## deviance   79.700   3.455  74.147 77.336 79.209 81.526 88.088 1.001  9000  
##  
## For each parameter, n.eff is a crude measure of effective sample size,  
## and Rhat is the potential scale reduction factor (at convergence, Rhat=1).  
##  
## DIC info (using the rule, pD = var(deviance)/2)  
## pD = 6.0 and DIC = 85.7  
## DIC is an estimate of expected predictive error (lower deviance is better).
```

```
traceplot(M6, mfrow=c(2, 3), ask=FALSE)
```



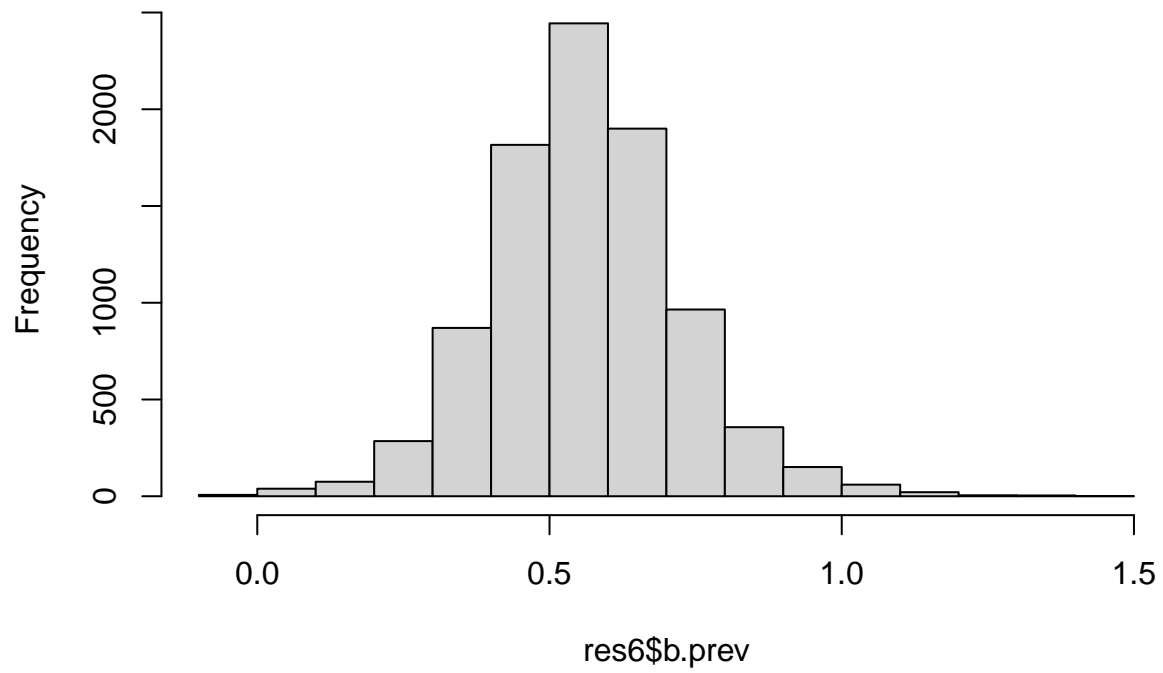


```
par(mfrow=c(1, 1))
```



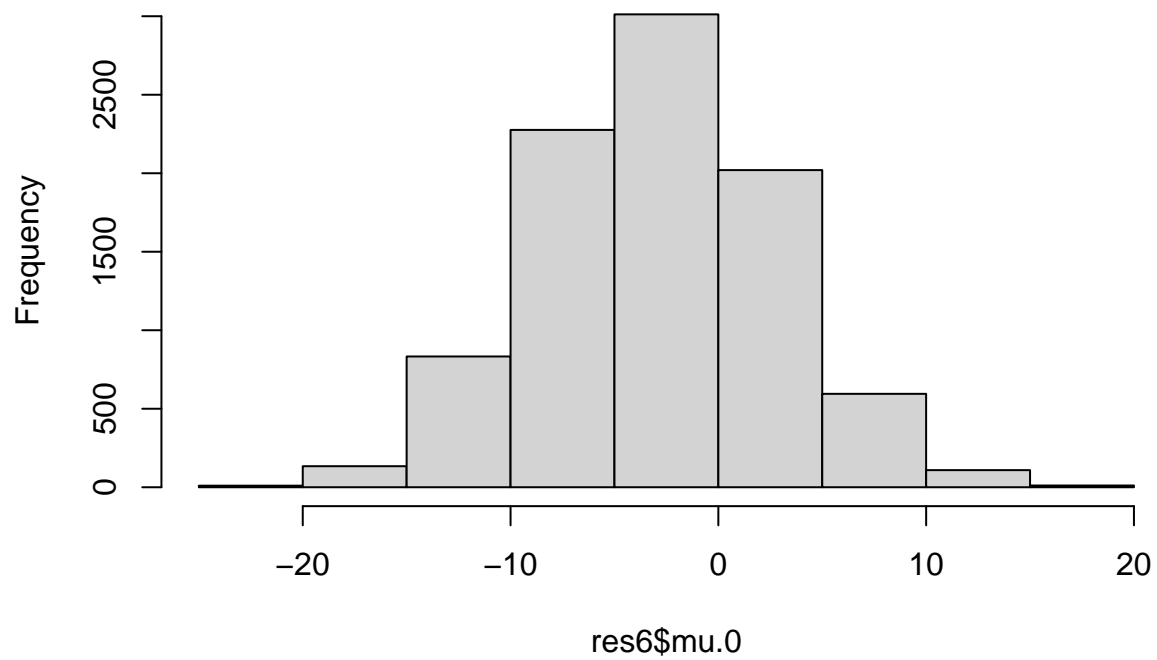
```
# On récupère les paramètres et on regarde leur distribution  
res6 <- M6$BUGSoutput$sims.matrix %>%  
  as.data.frame()  
hist(res6$b.prev)
```

**Histogram of res6\$b.prev**



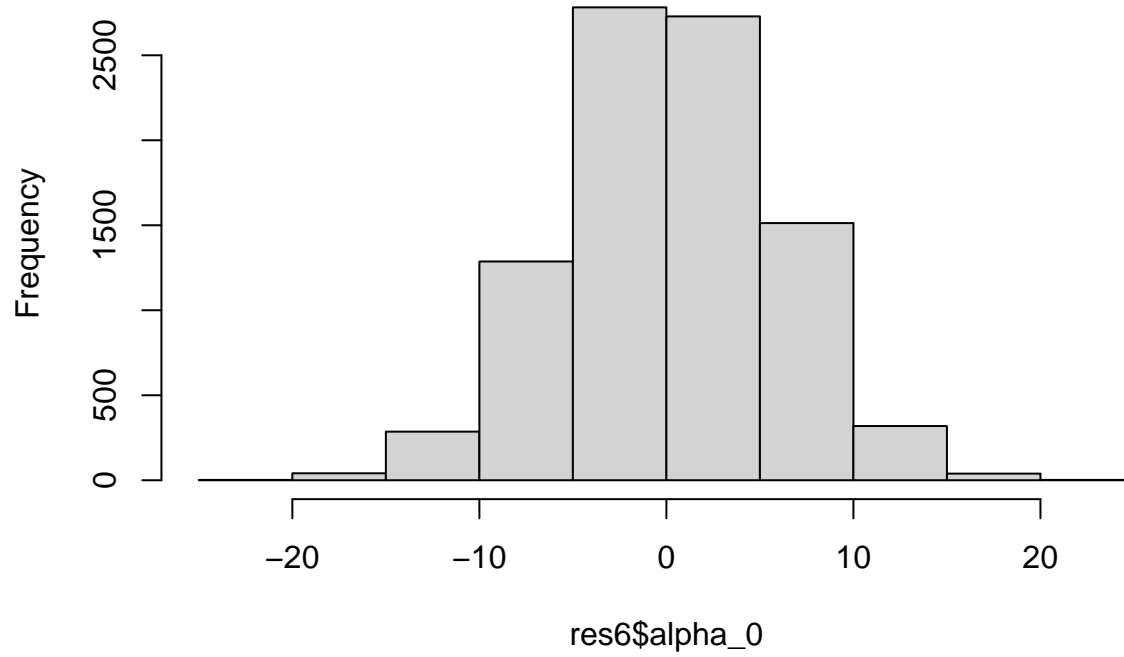
```
hist(res6$mu.0)
```

**Histogram of res6\$mu.0**



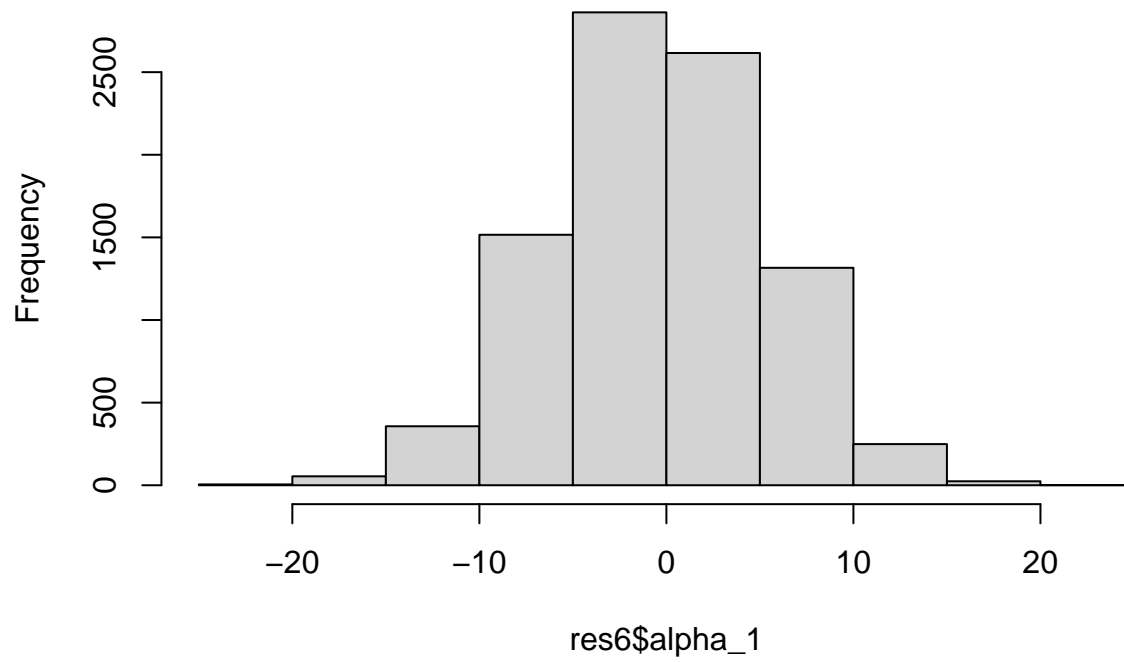
```
hist(res6$alpha_0)
```

**Histogram of res6\$alpha\_0**



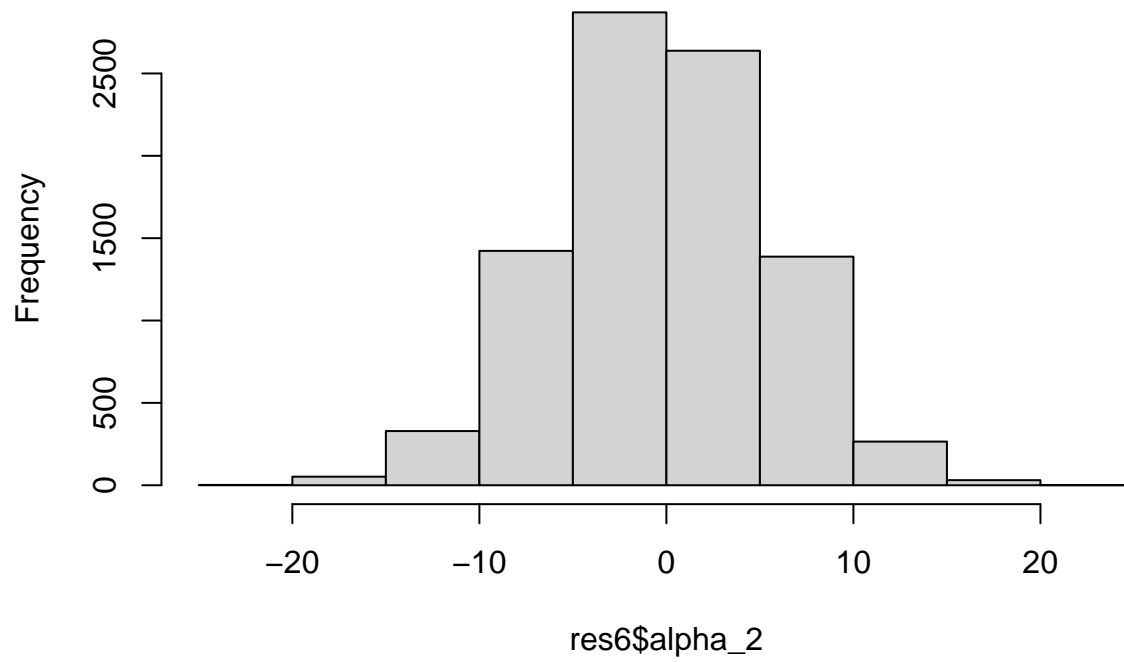
```
hist(res6$alpha_1)
```

**Histogram of res6\$alpha\_1**



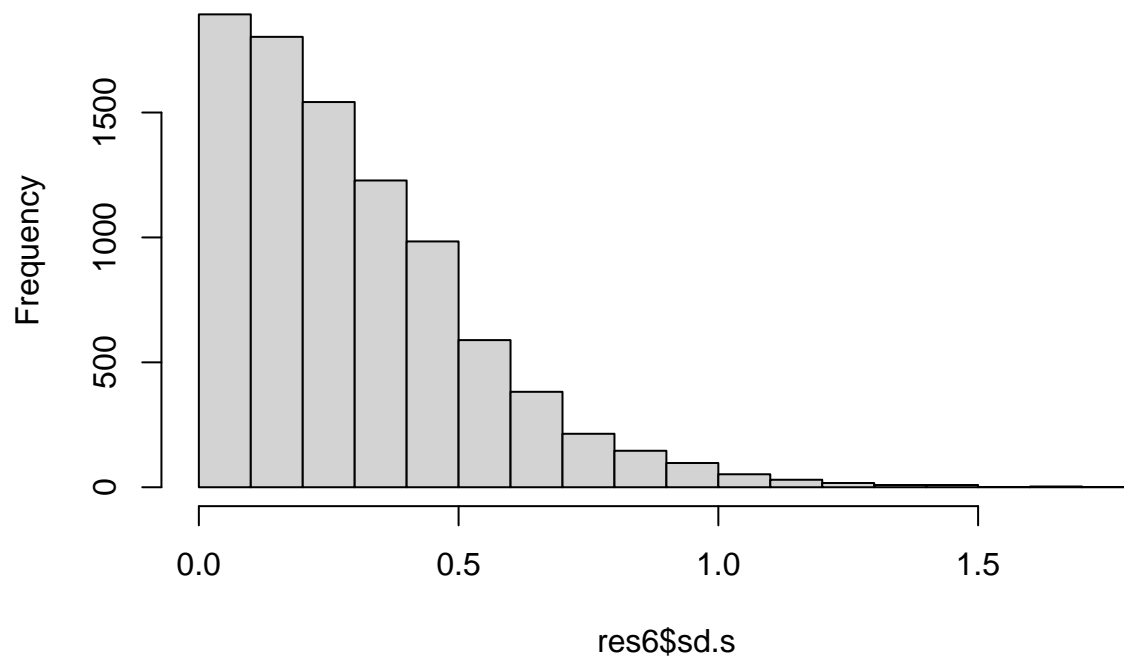
```
hist(res6$alpha_2)
```

**Histogram of res6\$alpha\_2**



```
hist(res6$sd.s)
```

## Histogram of res6\$sd.s

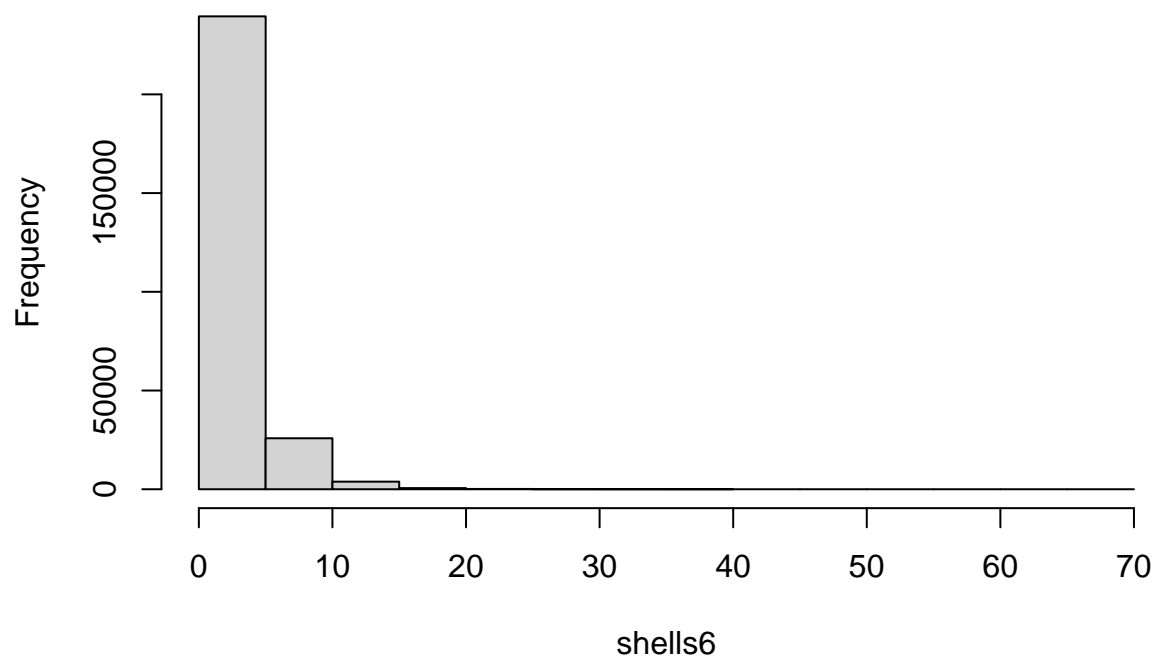


```
# Rétrotransformation
```

```
shells6 <- matrix(NA, ncol=nrow(gopher), nrow=nrow(res6))  
for (i in 1:nrow(gopher)){  
  shells6[,i] <- gopher$Area[i] * exp(res6$mu.0 + res6$alpha_0 * gopher$Cov_2004[i] + res6$alpha_1 * gopher$Cov_2004[i])  
}  
hist(shells6)
```



## Histogram of shells6



```
mean(shells6)
```

```
## [1] 2.03404
```

```
# Récupérons le DIC
```

```
DIC6 <- M6$BUGSoutput$DIC
```

```
|—test—|—test—| |  $\lambda$  |  $\mu$  |
```