

Journées AppliBUGS, 27 novembre 2008

# Le RJmcmc dans le cadre de la sélection de variables

Applications du module JUMP sous  
WinBUGS 14

Jérémy Piffady, thèse 3ème année

# Cas de la régression

- $Z$  = vecteur des observations
- $X$  = matrice des covariables
- $\beta$  = vecteur des coefficients
- $\varepsilon$  = vecteur des erreurs

- $Z = X \cdot \beta + \varepsilon$

Identifier les covariables pertinentes au sein du vecteur  $\beta$

# Ex: la Processionnaire du pin

## Tomassone, 1993

- $Z = \log(\text{nombre de nids par arbre})$
- $X_1 = \text{altitude (en m)}$
- $X_2 = \text{pente (en } ^\circ)$
- $X_3 = \text{nbe de pin par placette de 5 ares}$
- $X_4 = \text{hauteur de l'arbre au centre de la placette}$
- $X_5 = \text{diamètre de l'arbre au centre de la placette}$

# Ex: la ProceSSIONNAIRE du pin

## Tomassone, 1993

- $X_6$  = note de densité du peuplement
- $X_7$  = orientation de la placette (1: N, 2: S)
- $X_8$  = hauteur des arbres dominants (en m)
- $X_9$  = nombre de strates de la végétation
- $X_{10}$  = peuplement pur ou mélangé (1: pur, 2: mélangé)

# Ex: la Processionnaire du pin

## Tomassone, 1993

- Modèle constant

$$- Z = a_0 + \varepsilon$$

- Modèle complet

$$- Z = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 + \dots + a_{10} \cdot x_{10} + \varepsilon$$

- Au total, il y a  $2^{10} = 1028$  modèles possibles

# Méthodes classiques

- Test de tous les modèles, calcul des AIC
  - Temps
- Sélection automatique de variables
  - Forward, backward, stepwise
  - Les résultats sont souvent différents
- Quelles méthodes en bayésien?

# Reversible Jump mcmc

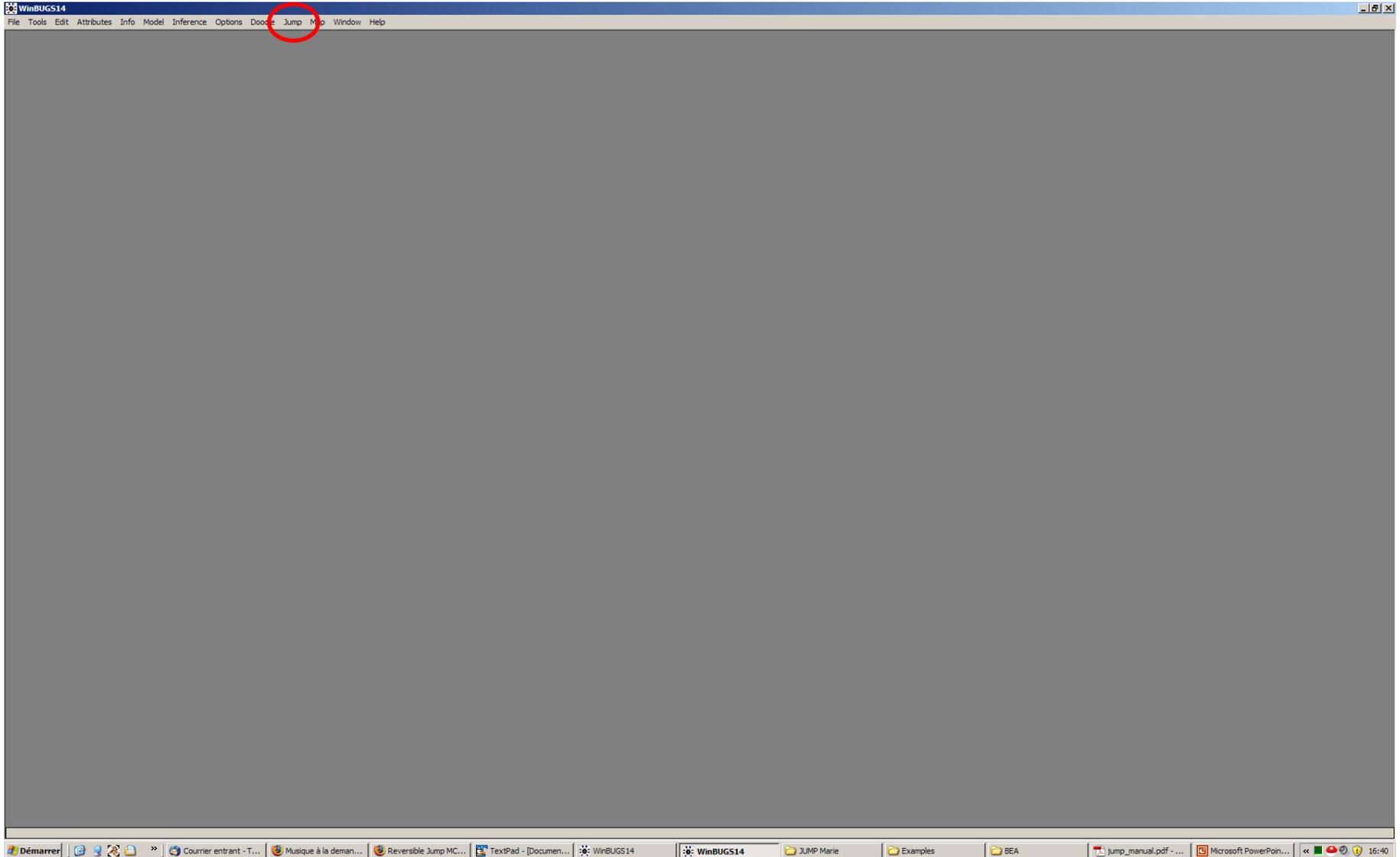
- Extension des algorithmes mcmc
- Exploration de l'espace des modèles
  - 1 itération = Nveau modèle + nulles valeurs de paramètres
- La règle d'acceptation nécessite l'écriture du rapport des crédibilités des modèles

# Procédure RJmcmc sous WinBUGS

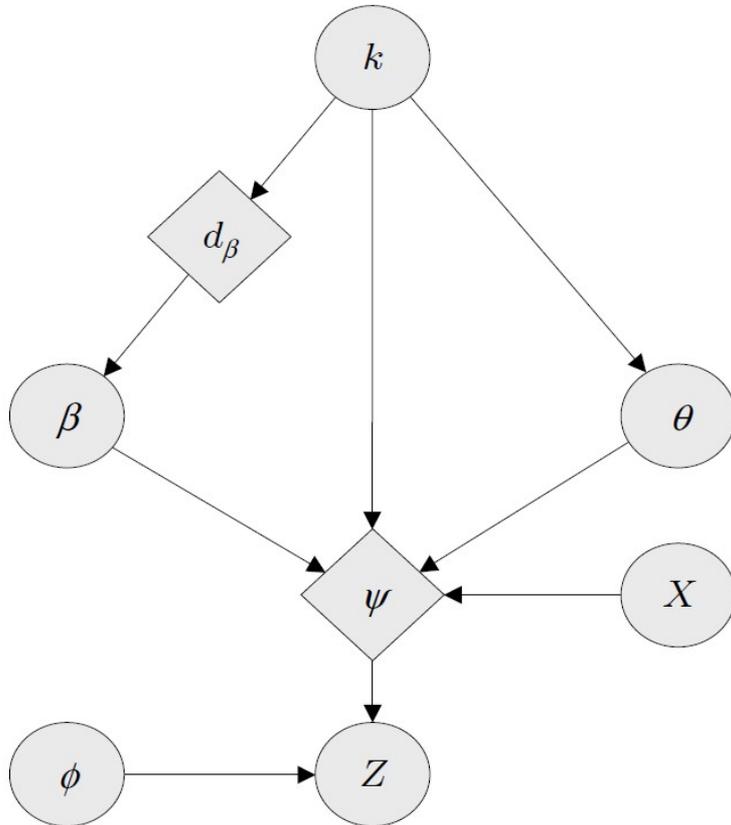
- WinBUGS JUMP interface
  - Lunn, Best et Whittaker (2006)
  - <http://www.winbugs-development.org.uk/rjmcmc>
- WinBUGS 1.4
  - Crash des autres versions

Lunn, D. J., Best, N. and Whittaker, J., 2005. Generic reversible jump MCMC using graphical models, *Technical Report EPH-2005-01*, Department of Epidemiology and Public Health, Imperial College London, UK.

# Aspect graphique



# Aspect graphique: doodle général



- $\psi = X \cdot \beta$ 
  - $k$  = nbe de paramètres
  - $\theta$  = vecteur binaire de structure
  - $\beta$  = vecteur des paramètres
  - $d_\beta$  = dimension du modèle
  - $X$  = matrice des covariables
- $Z$  = vecteur des observations
- $\Phi$  = vecteur des erreurs d'observations

# Formalisme en régression: fonctions utiles

- `Jump.lin.pred`
  - `psi[1:n] <- jump.lin.pred (X[1:n, 1:Q], k, beta.prec)`
    - Rappel :  $\psi = X.\beta$
  - $X$  = matrice des covariables
  - $k$  = variable aléatoire donnant le nbe de paramètres
  - `beta.prec` = paramètre de précision du prior sur les  $\beta$
- `Jump.lin.id`
  - `id <- jump.lin.id (psi[1:n])`
    - Stocke les structures successives lors des itérations

# Mises à jour

- « Partie fixe » du modèle
  - Échantillonneur de Gibbs
- Proposition d'un modèle
  - $\theta'$  = vecteur binaire de structure
  - $d_{\beta}'$
  - $\beta'$  = vecteur des paramètres

Processus de naissance et de mort des variables

- Garantie de la réversibilité

# Internalisation de $\beta$ et $\theta$

- $\theta$ 
  - jump.lin.id : calculs automatiques via JUMP
- $\beta$ 
  - pas de contrôle direct du prior.
  - normale multivariée: choix logique / conjugaison
  - variance à définir : la même pour tous

# Retour de la Processionnaire du pin

```
model {  
  Y[1:n] <- jump.lin.pred(X[1:n, 1:Q], k, beta.prec)  
  id <- jump.model.id(Y[1:n])  
  k ~ dbin(0.5, Q)  
  beta.prec ~ dgamma(0.001, 0.001)  
  
  for (i in 1: n) {  
    Z[i] ~ dnorm(Y[i], tau)  
  }  
  tau ~ dgamma(0.001, 0.001)  
}
```

# Retour de la Processionnaire du pin

The screenshot shows the WinBUGS 1.4 interface with the following components:

- Main Window:** Displays the following R code:

```
model {
tau ~ dgamma(0.001, 0.001)
for (i in 1:n) {
Z[i] ~ dnorm(Y[i], tau)
}
Y[1:n] <- jump.lin.pred(X[1:n, 1:Q], k, beta.prec)
id <- jump.model.id(Y[1:n])
k ~ dbin(0.5, Q)
beta.prec ~ dgamma(0.001, 0.001)
}

tau = 1,
beta.prec = 1,
k = 10

list(
n = 33,
Q = 10,
Z = c(0.862889955, 0.385262401, 0.122217633, -0.162518929, -1.427116356, 0.39877612, -1.203972804,
-2.659260037, 1.096612289, 0.19062036, -0.967584026, -0.356674944, 0.970778917, 0.717639793, 0.559615788, -2.813410717,
-2.140220829, 0,
-0.891598119, -0.328504067, -0.400477567, -2.120263596, -0.030459207, -2.859260037, -2.302585093, -0.385662461, -2.040220829,
1.608437912, 0.086177696, -1.714798428, -1.049822124, -1.560647748, -3.506557697),
X = structure(Data=c(
-0.968992248, -1, -1, -0.5, -0.11627907, -1.42857143, -8, -0.666666667, -1, -1.333333333,
0.131782946, -0.142857143, -0.3, -0.1, 0.627906977, -0.428571429, -1, -0.458333333, -0.5, -0.333333333,
-0.728682171, -0.142857143, -0.6, -2, 1, -1.744186047, -0.714285714, -0.5, -1.333333333, -0.833333333, -1.333333333,
-0.550387597, -0.142857143, 0.7, -1.5, -1.418604651, 0.714285714, 0, -0.25, 0.5, -0.666666667,
0.248062016, 0.428571429, -0.4, -0.8, -1.069767442, -0.571428571, 0, -0.375, -0.333333333, -1.666666667,
-0.581395349, -0.285714286, -1, -0.1, -0.11627907, -1.42857143, 0, -0.708333333, -1.166666667, -1.333333333,
0.751937984, 1.42857143, 1, -1.5, -1.674418605, 1.285714286, 1, 0.333333333, 0.333333333, 0.666666667,
-0.124031008, 2.428571429, -0.4, 1.2, 1, -0.428571429, -2, 0.125, -0.333333333, -0.666666667,
-1.534883721, -0.714285714, -0.9, -1, -0.627906977, -1.42857143, 0, -0.833333333, -0.833333333, 0.666666667,
-1.937984496, 0.714285714, -0.2, -0.2, -0.76744186, 0.285714286, 0.5, -0.291666667, 0.666666667,
-1.23255814, -0.714285714, 0.6, 1, 0.325581395, 0.857142857, -1, 0.666666667, 1.5, -0.333333333,
-1.108527102, 1.714285714, 2, 1.0, 0.9, 1.465116279, 2.42857143, -1.5, 1.583333333, 1.333333333, 0.666666667,
-1.13782946, -2, -1, -1.3, -1.11627907, -1.42857143, 0, -1.458333333, -1.5, -0.666666667,
-0.534883721, -1.42857143, -1, 1.0, 0.6, 0.976744186, -1.42857143, 0.5, -0.708333333, -1.5, -1.333333333,
-0.573643411, -0.428571429, -0.9, -0.3, 0.255813953, -1, 0, -0.541666667, -1.166666667, 0,
1.635658915, 1.285714286, 2, 1.2, 0.581395349, 1.857142857, 1, 0.825, 1.333333333, 0.333333333,
1.775193798, -0.285714286, 0.8, 1.1, 1.19354884, 1.428571429, 1.0, 708333333, 1.166666667, -1.666666667,
-0.15503876, 0.42857143, -0.5, -0.7, 0.093023256, -0.571428571, -2.5, -0.125, 0.166666667, 0.333333333,
-0.069767442, 0.714285714, -0.3, -1.4, -0.906976744, -0.428571429, 0.5, -1.041666667, 0.666666667, 0.666666667,
0.790897674, 1.428571429, 0.0, 0.1, -0.02255814, 0.857142857, -0.5, 0.666666667, 0.666666667, 0.333333333,
1.937984496, -1.285714286, 2, 1.2, 0.581395349, 1.857142857, 1, 0.825, 1.333333333, 0.333333333,
0.558139535, -0.428571429, 0.5, -0.3, -0.11627907, 0.571428571, -0.5, 0.833333333, 0.333333333, 0,
0.403100775, -0.7, 0.8, 1.046511628, -0.857142857, 0.5, -0.291666667, 0.666666667, 0.333333333,
1.930232558, -0.714285714, 2.0, 7.0, 581395349, 0.857142857, 0.5, 0.125, 0.333333333, 0.666666667,
0.550387597, 2.285714286, 0.2, 0.2, -0.02255814, -0.142857143, -0.5, 0.125, 0.166666667, -1.333333333,
0.52713783, -0.285714286, -0.8, 0.2, 0.697674419, -0.857142857, 0, -0.5, 0.666666667,
0.837209302, -0.857142857, 0.7, 2.1, 325581395, 1.285714286, 0.5, 2.583333333, 1.166666667, -1.666666667,
0.186046512, -0.714285714, -1.8, -2.209302326, -1.42857143, 0, -1.625, -1.166666667, 0,
-0.906976744, -0.857142857, -0.9, -1, -0.88372093, -1, 0, -0.875, -1.166666667, 0.666666667,
-0.984496124, -0.142857143, 0.4, -0.6, -0.930232558, 0.285714286, -0.5, -0.041666667, 1.333333333, 0.666666667,
```
- Specification Tool:** Shows 'check model' and 'load data' buttons, and a 'num of chains' set to 1.
- Sample Monitor Tool:** Shows 'node' dropdown, 'chains' set to 1, 'to' set to 1, 'percentiles' set to 2.5, 5, 10, 25, median, 75, 90, 95, 97.5.
- Update Tool:** Shows 'updates' set to 10000, 'refresh' set to 1000, 'update' button, 'thin' set to 1, 'iteration' set to 10000, and checkboxes for 'over relax' and 'adapting'.
- Jump Summary Tool:** Shows 'id node' dropdown, 'history' button, 'beg' set to 1, 'end' set to 100000, 'razor' set to 0.99, and 'mixing' and 'table' buttons.

posterior prob.	cumulative prob.
0.027324	0.027324
0.022628	0.049952
0.015084	0.065036
0.014854	0.07989
0.013898	0.093788
0.013326	0.107114
0.013274	0.120388
0.011588	0.131976
0.010328	0.142304
0.009748	0.152052
0.009438	0.16149
0.00941	0.1709
0.009094	0.179984
0.008434	0.188418
0.008376	0.196794
0.008336	0.20513
0.008156	0.213288
0.007824	0.22111
0.007448	0.228558
0.00704	0.235588
0.007024	0.242622
0.006704	0.249326
0.006616	0.255942
0.006476	0.262418
0.006314	0.268732
0.006194	0.274926
0.00619	0.281116
0.00591	0.287026
0.005826	0.292852
0.005308	0.29816
0.00503	0.30319
0.004938	0.308128
0.004934	0.313062
0.004852	0.317914
0.004716	0.32283
0.00459	0.32772
0.00455	0.33177
0.00454	0.33631
0.004494	0.340804
0.004292	0.345096
0.004276	0.349372
0.004266	0.353638
0.004188	0.357826
0.00418	0.362006
0.004168	0.366174
0.00416	0.370334
0.004086	0.37442
0.003978	0.378398
0.003956	0.382354
0.003956	0.38631
0.003944	0.390254
0.003796	0.39405
0.003776	0.397826
0.00347	0.401296
0.003404	0.4047
0.003404	0.408104
0.003386	0.41149
0.003374	0.414864
0.003368	0.418232
0.00331	0.421542
0.003288	0.42483
0.00328	0.42811
0.003214	0.431324
0.003182	0.434506
0.003178	0.437684
0.003138	0.440822
0.003114	0.443936

check model   num of chains 1

node [ ] chains 1 to 1 percentiles 2.5 5 10 25 75 90 95 97.5

beg 1 end 1000000 thin 1

clear  trace history density

stats  quantiles bgr diag auto cor

load inits for chain 1

updates 500000 refresh 1000

update thin 1 iteration 500000

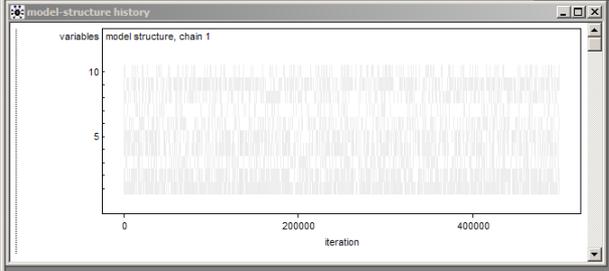
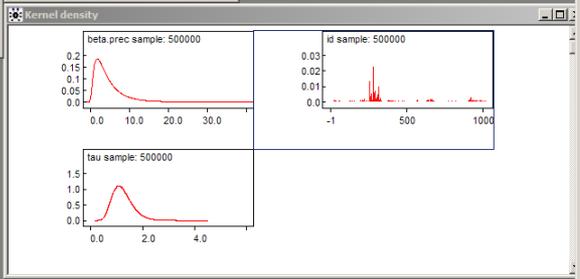
over relax  adapting

**Jump Summary Tool**

id node: id history

beg: 1

end: 1000000 razor: 0.99



lu

# Résultats

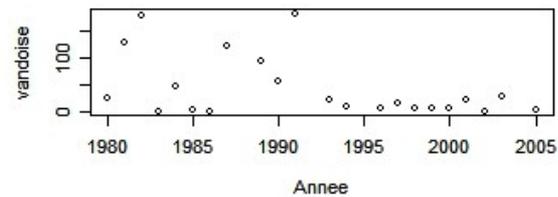
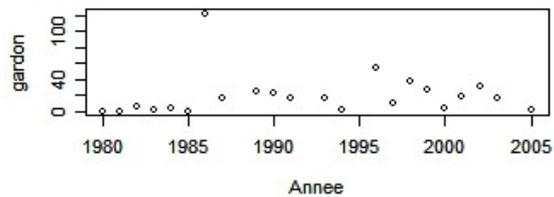
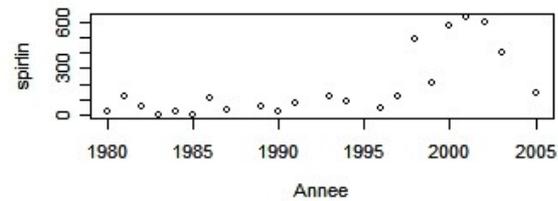
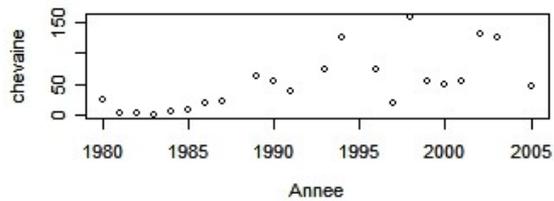
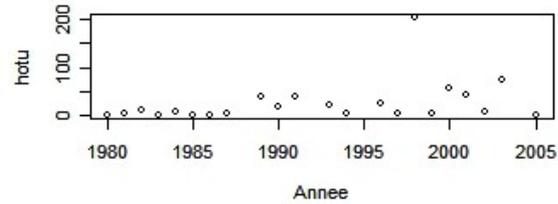
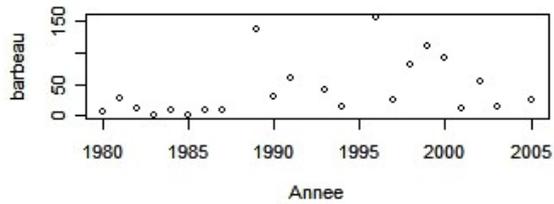
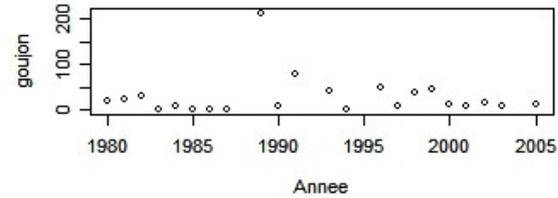
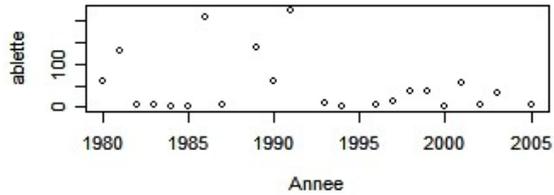
- Stepwise classique
  - $X_1, X_2, X_4$  et  $X_5$
- Onglet jump
  - Table

model structure	posterior prob.	cumulative prob.	AIC stepwise
1101100000	0.02755166667	0.02755166667	84.724
1101100010	0.02180033333	0.049352	85.717
1101110010	0.01462366667	0.06397566667	85.978

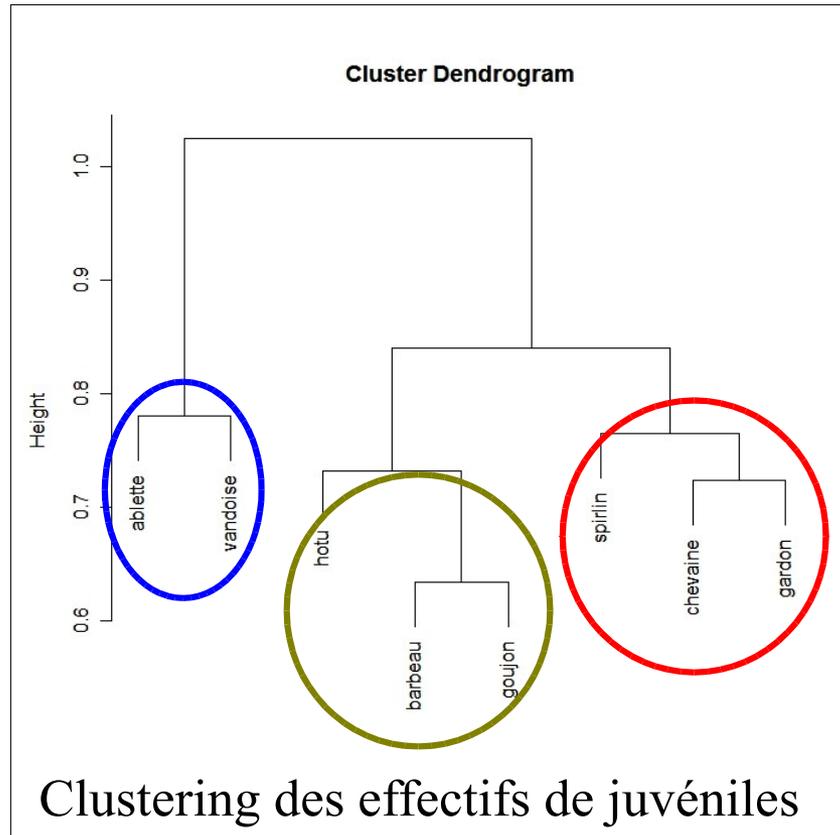
# Réponse des peuplements piscicoles juvéniles aux variations de régimes thermiques et hydriques

- Etude monosite
- 26 ans de données de captures par pêche électrique (octobre – novembre)
- Juvéniles de l'année (repro mars – avril)
- Suivis quotidiens de température et de débit

# Variabilité biologique



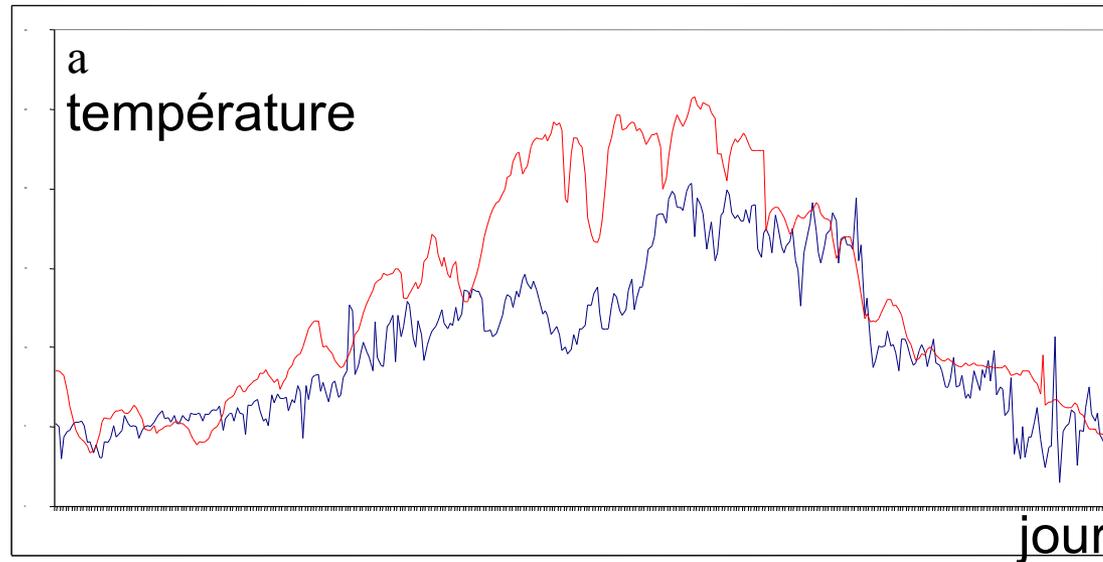
# Variabilité biologique



- Tau de kendall
- Distance =  $1 - \tau / 2$

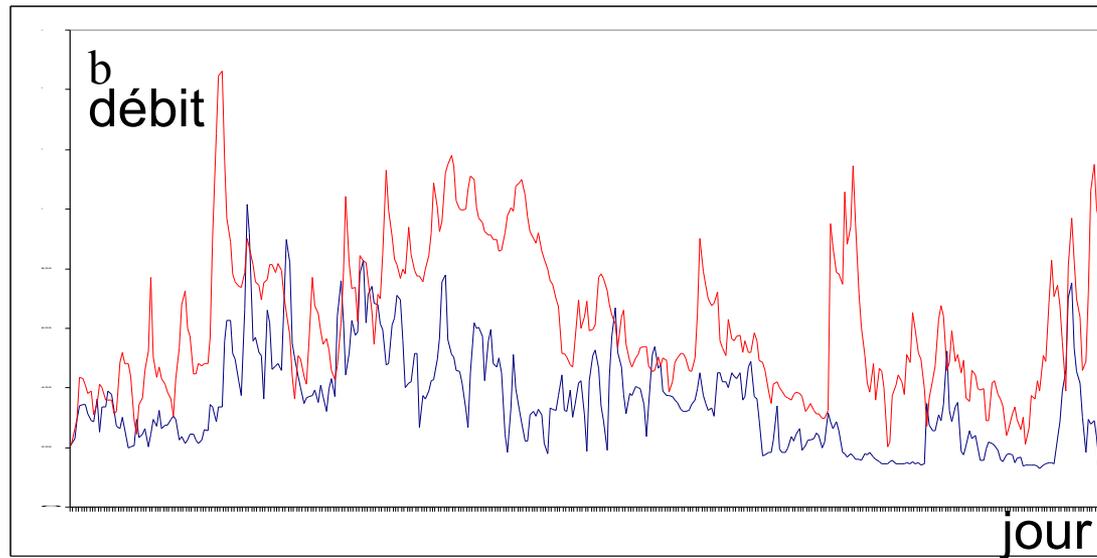
# Variabilité environnementale

- température



# Variabilité environnementale

- Débits



# Covariables pressenties

- Etude de l'influence des régimes thermique et hydrique sur le nombre de juvéniles.
  - $X_1$  = jour d'atteinte du seuil 12°C
  - $X_2$  = jour d'atteinte du max de température
  - $X_3$  = chaleur relative au printemps
  - $X_4$  = chaleur relative en été
  - $X_5$  = débit moyen au printemps
  - $X_6$  = débit moyen en été
  - $X_7$  = débit max en été

# Structure du modèle statistique: exemple d'un groupe à 2 espèces

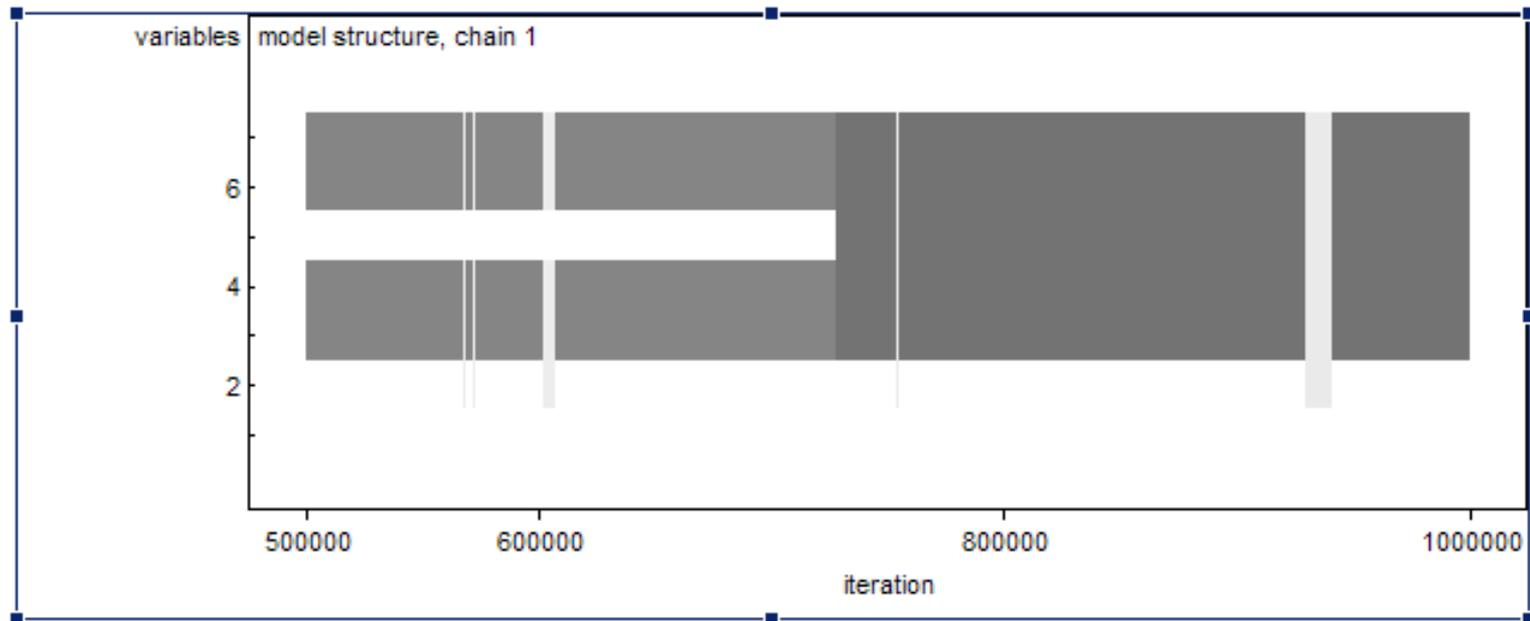
- Réponse à l'environnement au niveau du groupe
- Effectifs relatifs
- $\psi_{1j} = X \cdot \beta = \log(\text{effectif}_{1j})$
- $\psi_{2j} \sim N(\psi_{1j}, 10)$
- $Z_{i,j} =$  observation poissonnienne de la densité
  - $Z_{i,j} \sim P(\exp(\psi_{ij}))$

# Code WinBUGS pour un groupe à 2 espèces

```
model{  
psi1[1:n] <- jump.lin.pred(X1[1:n, 1:Q], k1 , beta.prec1)  
id1 <- jump.model.id(psi1[1:n])  
k1 ~ dbin(0.5, Q)  
beta.prec1 <- 0.001  
for (e in 1: n) {  
    psitmp[e] ~ dnorm(psi1[e], 10000)  
    psitmp2[e] ~ dnorm(psitmp[e], 0.1)  
    Y1[e] <- exp(psitmp[e])  
    Z1[e] ~ dpois(true_recruABL[e])  
    Y2[e] <- exp(psitmp2[e])  
    Z2[e] ~ dpois(true_recruVAN[e])  
    }  
}
```

# Résultats

- History



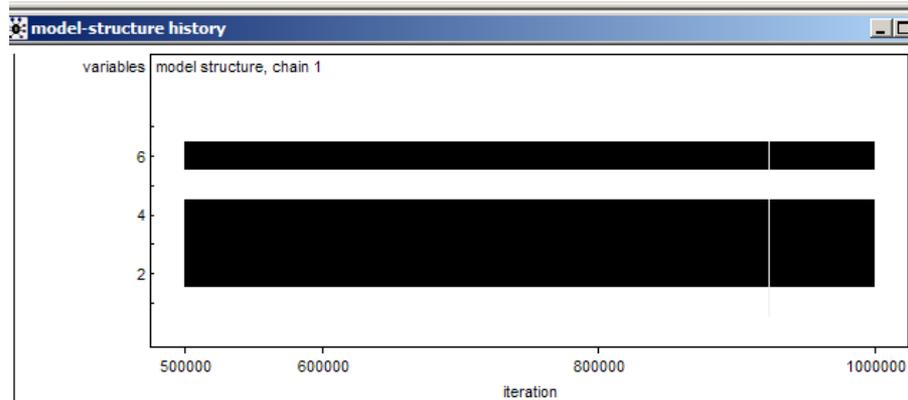
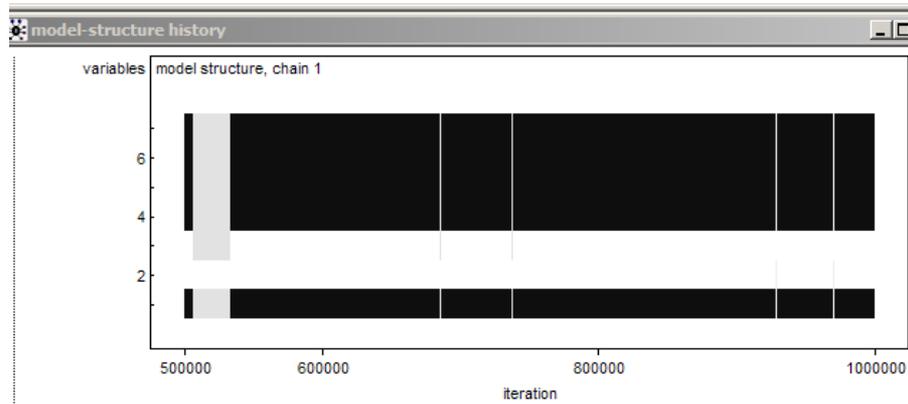
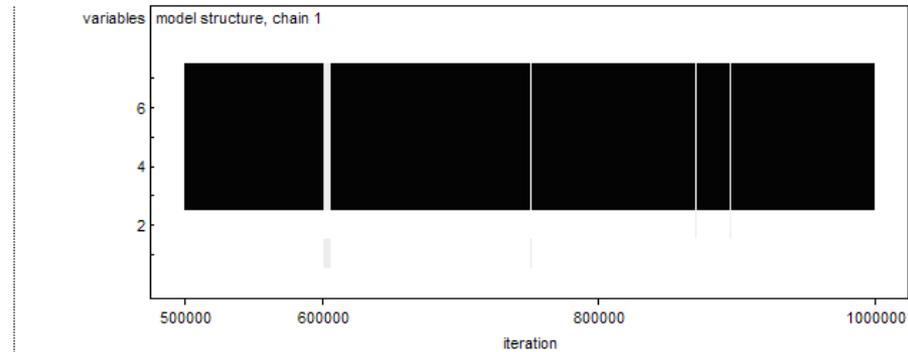
# Résultats

- table

model structure	posterior prob.	cumulative prob.
0011111	0.519404	0.519404
0011011	0.442196	0.9616
0111111	0.023976	0.985576
0111011	0.012416	0.997992

variable no.	marginal prob.
1	0.002008
2	0.036392
3	1.0
4	1.0
5	0.54506
6	1.0
7	1.0

# Résultats: 3 groupes



To be continued...