

Buzzing - Rapport d'avancement

Sommaire

1 Modélisation du buzzing chez les narvals sans exposition	2
1.1 Estimation de l'effet de la profondeur	2
1.2 Estimation de l'autocorrélation dans l'émission de buzz	2
1.2.1 Recherche de la mémoire optimale	2
1.2.2 Note sur le temps d'ajustement des modèles	4
1.2.3 Régression bi-exponentielle sur les coefficients autorégressifs	5
1.2.4 Coefficients de profondeur	5
2 Estimation de l'effet de l'exposition sur le buzzing	6
2.1 Modélisation	6
2.2 Intervalles de confiance des coefficients d'exposition	7
2.2.1 Variation des coefficients selon des lois normales univariées	7
2.2.2 Variation des coefficients selon une loi normale multivariée	9

Nous souhaitons observer chez des narvals l'effet de l'exposition à des perturbations humaines sur leur capacité à se nourrir. Lorsqu'elles se nourrissent, ces baleines émettent des sons spécifiques (buzz). Nous nous intéresserons donc à la fréquence d'émission de ces buzz (buzz/min), ce qui peut être modélisé par un modèle de Poisson.

1 Modélisation du buzzing chez les narvals sans exposition

1.1 Estimation de l'effet de la profondeur

Pour se nourrir, les narvals doivent plonger profondément alors que le reste du temps, ils restent “proche” de la surface. Il faut donc inclure au modèle la profondeur à laquelle se trouvent les baleines quand elles émettent ou non des buzz. La relation entre l'émission de buzz et la profondeur n'étant pas linéaire, la profondeur a été remplacée par une variable explicative la décrivant par un polynôme de degré 3.

1.2 Estimation de l'autocorrélation dans l'émission de buzz

1.2.1 Recherche de la mémoire optimale

L'émission d'un buzz à un instant t est corrélé à l'émission ou non de buzz aux instants précédents t . Cet effet mémoire doit donc être ajouté au modèle et pour cela nous devons déterminer la mémoire maximum qu'il faut autoriser au modèle.

Afin de trouver la mémoire optimale, nous avons utilisé la démarche suivante :

1. $from.ml = 1 ; to.ml = N$
2. Tant que ($to.ml - from.ml > 2$) :
 1. Pour $ml_i^k = from.ml + (i - 1) * \lfloor \frac{to.ml - from.ml}{M-1} \rfloor$, $i = 1...M$:
 1. ajustement d'un modèle de Poisson avec ml_i^k éléments mémoire
 2. calcul du BIC
 2. $i_{opt} = argmin BIC_i$
 3. $ml_{opt} = ml_{i_{opt}}^k$
 4. $from.ml = ml_{i_{opt}-1}^k ; to.ml = ml_{i_{opt}+1}^k$

1.2.1.1 Sans effets aléatoires Dans un premier temps, nous n'avons pas inclus d'effets aléatoires sur les individus pour estimer la mémoire optimale :

$$Buzz \sim Ind + spline(Depth) + Lag_1 + \dots + Lag_N$$

Nous avons fixé $N = 300$ et $M = 10$.

Nous obtenons une mémoire optimale de 60 secondes. La figure 1 permet de voir que cet optimal semble bien correspondre à un minimum global.

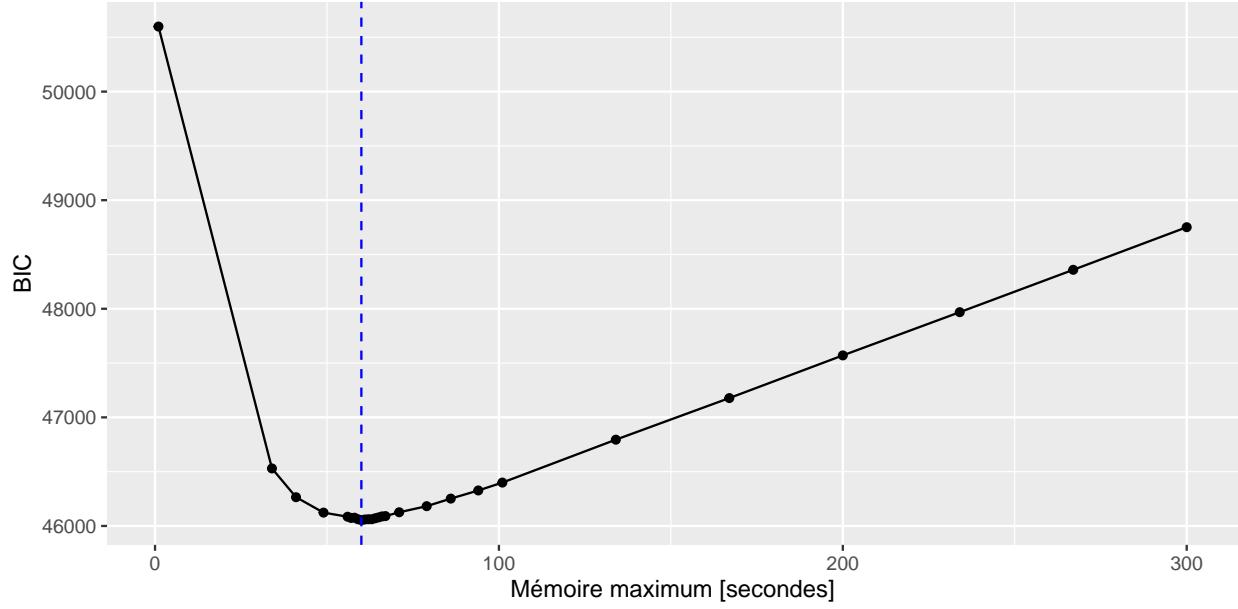


Figure 1: BIC en fonction de la mémoire maximum

1.2.1.2 Avec effets aléatoires Nous avons plusieurs observations par individu, ce qui implique qu'elles ne sont pas indépendantes. Pour prendre en compte cette dépendance, nous avons tenté d'inclure un effet aléatoire sur les individus :

$$Buzz \sim (1|Ind) + \text{spline}(Depth) + Lag_1 + \dots + Lag_N$$

Compte tenu du résultat obtenu précédemment, nous avons tout de suite restreint l'intervalle de recherche initial à [30, 90].

La figure 2 présente les résultats obtenus. La courbe est similaire à celle obtenue sans l'ajout d'effets aléatoires.

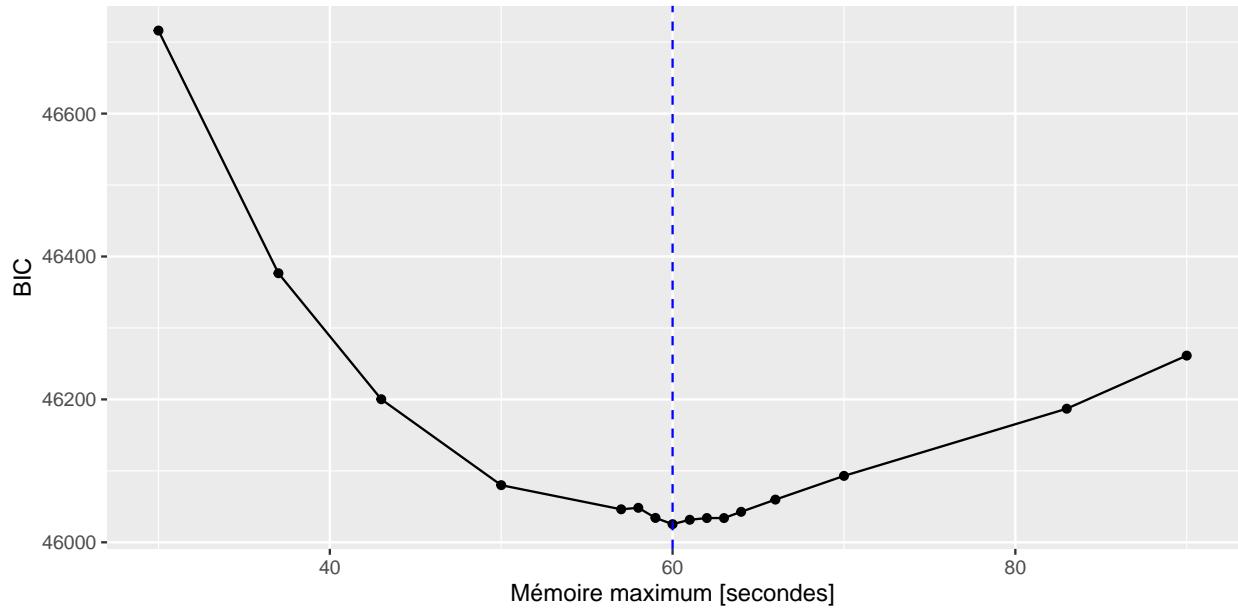


Figure 2: BIC en fonction de la mémoire maximum - modèle mixte

1.2.2 Note sur le temps d'ajustement des modèles

Le temps d'ajustement des modèles, et en particulier des modèles mixtes, augmente fortement lorsque que le nombre de paramètres à ajuster augmentent.

Comme nous pouvons le voir sur la figure 3 cette augmentation est linéaire pour les modèles classiques, alors que pour les modèles mixtes celle-ci est quadratique.

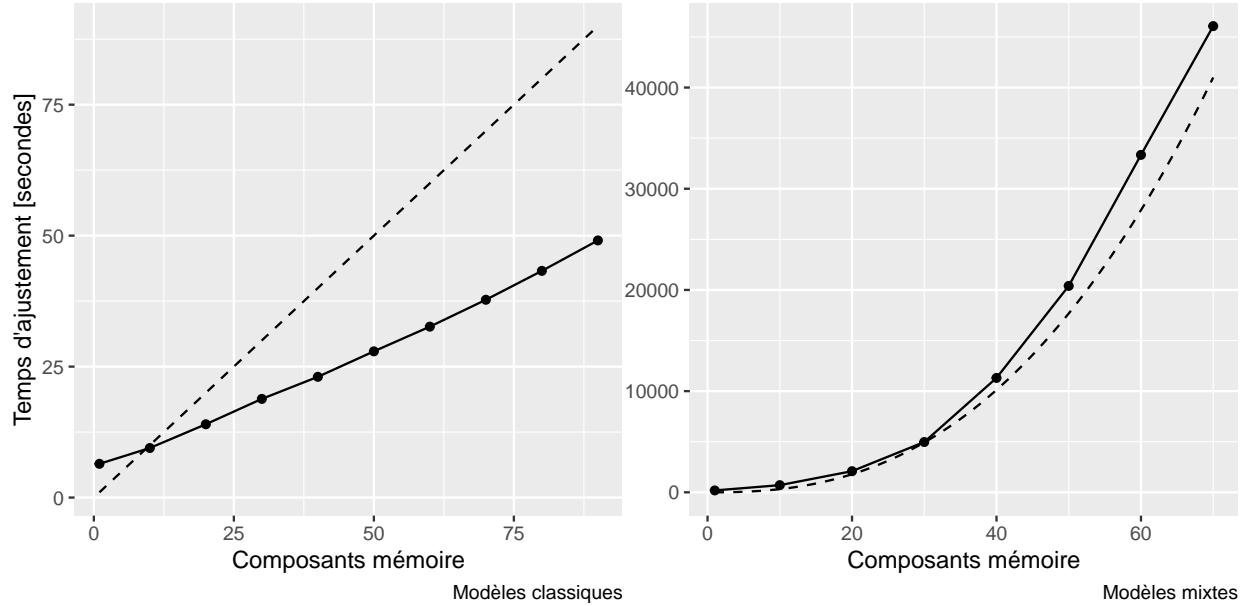


Figure 3: Temps d'ajustement (en secondes) des modèles en fonction de la mémoire maximum

1.2.3 Régression bi-exponentielle sur les coefficients autorégressifs

Nous avons vu que les modèles mixtes sont bien plus longs à ajuster que les modèles classiques, aussi, afin de réduire le nombre de variables et le temps d'ajustement, nous appliquons une régression bi-exponentielle aux coefficients auto-régressifs obtenus précédemment. Ainsi nous passons de 60 à 4 variables.

La figure 4 illustre les composantes de la mémoire ajustées pour un décalage maximum de 60. Les points sont les contributions individuelles de chaque décalage, la courbe correspond à la régression bi-exponentielle.

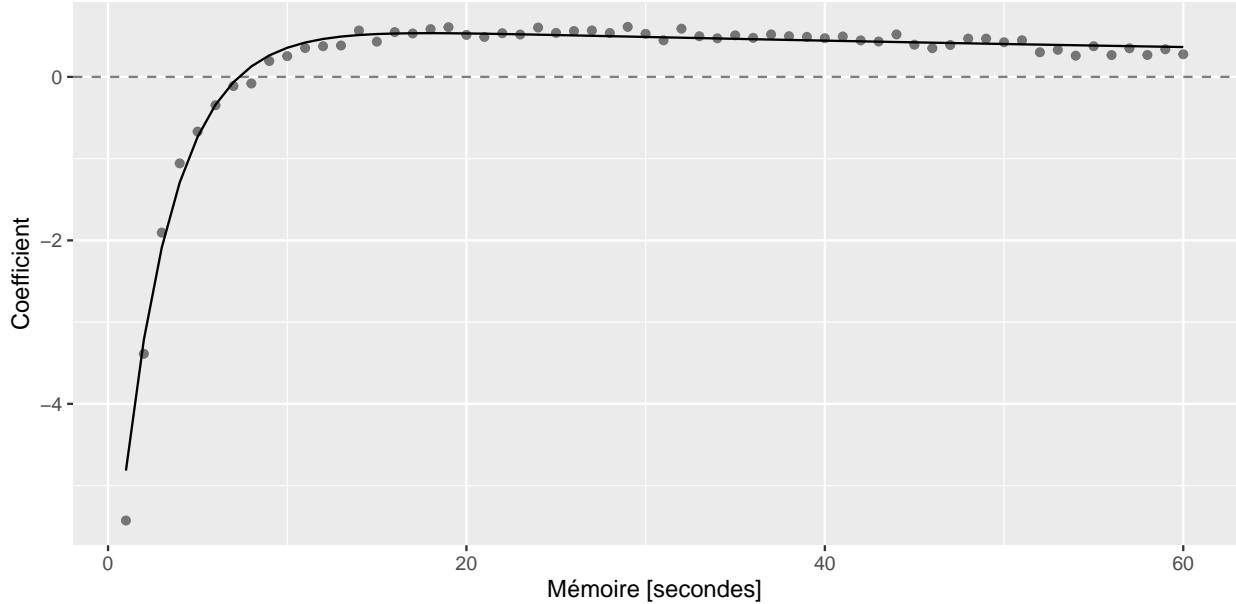


Figure 4: Régression bi-exponentielle des coefficients auto-régressifs

Les coefficients obtenus suite à la régression bi-exponentielle sont présentés dans la table 1.

Table 1: Coefficients autorégressifs obtenus par régression bi-exponentielle

term	estimate	std.error	statistic	p.value
A1	-7.7413301	0.3287556	-23.54737	0
lrc1	-1.0545001	0.0407916	-25.85088	0
A2	0.6544679	0.0427976	15.29218	0
lrc2	-4.6329847	0.1880983	-24.63066	0

1.2.4 Coefficients de profondeur

Les coefficients de profondeur du modèle linéaire mixte sans exposition sont proposés table 2.

Table 2: Coefficients de profondeur

term	estimate	std.error	statistic	p.value
splineDepth1	0.0265527	0.1308616	0.2029069	0.8392078
splineDepth2	-0.1254979	0.2069123	-0.6065272	0.5441647
splineDepth3	-23.4762851	2.4222863	-9.6917878	0.0000000
splineDepth4	-32.6394721	3.3383763	-9.7770500	0.0000000

2 Estimation de l'effet de l'exposition sur le buzzing

2.1 Modélisation

L'exposition aux perturbations est exprimée par $1/dist$ où $dist$ est la distance en kilomètres séparant l'animal du bateau dont émane la perturbation.

Nous avons ajusté le modèle mixte suivant :

$$Buzz \sim (1|Ind) + offset(ARDepth) + spline(Expo)$$

Les coefficients estimés par le modèle mixte sans exposition sont fixés dans celui-ci afin que les coefficients ajustés pour la variable d'exposition s'interprètent comme un effet par rapport au comportement normal (sans exposition). De plus, cela permet de réduire sensiblement le coût calculatoire d'ajustement.

Table 3: Coefficients du modèle mixte incluant l'exposition

effect	group	term	estimate	std.error	statistic	p.value
fixed	NA	(Intercept)	-4.5851696	0.0350968	-130.6436	0
fixed	NA	ns(X, knots = quantile(data\$X[data\$X > 0], 1:2/3))1	-1.1458123	0.0021544	-531.8501	0
fixed	NA	ns(X, knots = quantile(data\$X[data\$X > 0], 1:2/3))2	-58.8723530	0.0543014	-1084.1782	0
fixed	NA	ns(X, knots = quantile(data\$X[data\$X > 0], 1:2/3))3	-112.7904686	0.1069174	-1054.9306	0
ran_pars	Ind	sd_(Intercept)	0.0859693	NA	NA	NA

Table 4: Statistiques du modèle mixte incluant l'exposition

sigma	logLik	AIC	BIC	deviance	df.residual
1	-34393734001	68787468012	68787468075	51675872055	2367453

La figure 5 présente plusieurs graphiques utiles pour valider visuellement le modèle : nous sommes satisfaits de l'absence de corrélation entre les résidus constatée sur les deux premiers et du comportement gaussien visible sur le dernier.

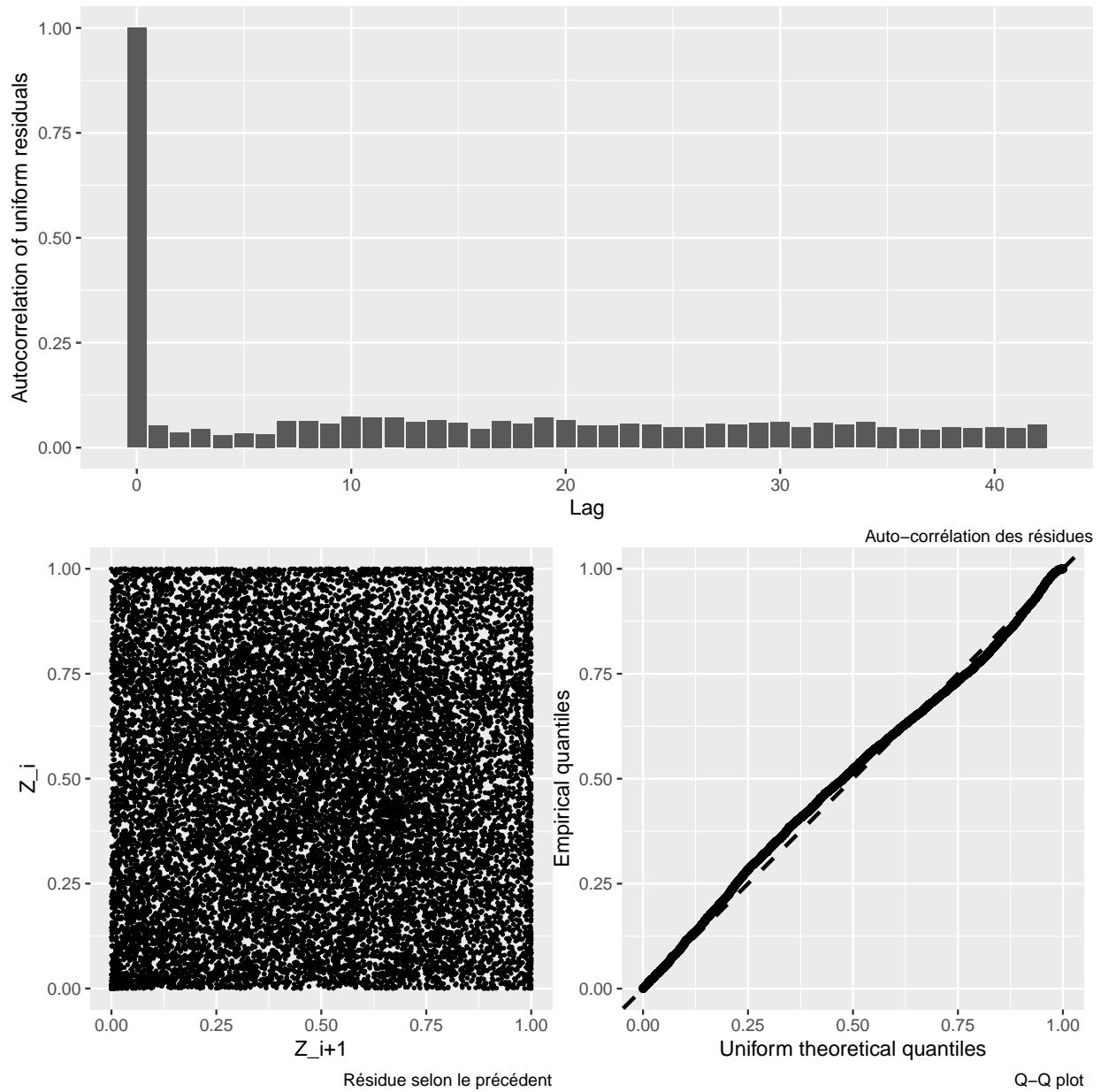


Figure 5: Validation graphique du modèle

2.2 Intervalles de confiance des coefficients d'exposition

2.2.1 Variation des coefficients selon des lois normales univariées

A expliquer.

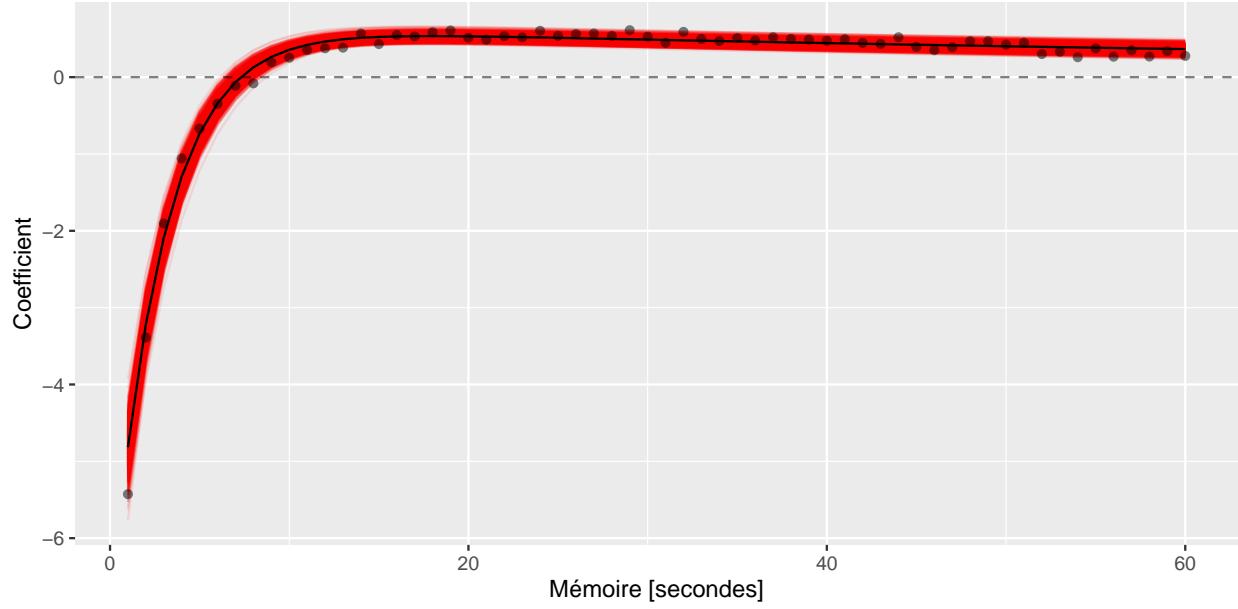


Figure 6: Variation des coefficients autorégressif selon des lois normales univariées

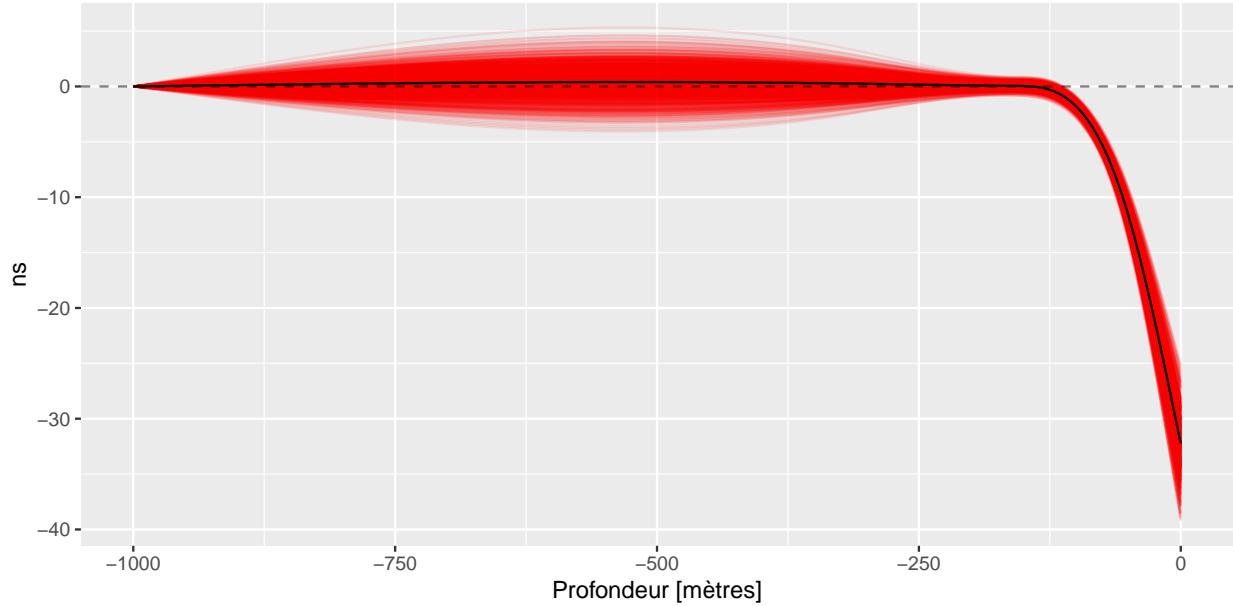


Figure 7: Variation des coefficients de profondeur selon des lois normales univariées

Table 5: Intervalles de confiance dans le cas de normales univariées

	MC - lower	MC - upper	SE - lower	SE - upper
(Intercept)	-6.441511	-2.968321	-4.638673	-4.494102
ns(X, knots = quantile(data\$X[data\$X > 0], 1:2/3))1	-11.898716	11.294079	-1.430436	-1.422075
ns(X, knots = quantile(data\$X[data\$X > 0], 1:2/3))2	-152.959472	1.047762	-56.014702	-55.803478
ns(X, knots = quantile(data\$X[data\$X > 0], 1:2/3))3	-306.428946	11.706679	-106.846662	-106.430830

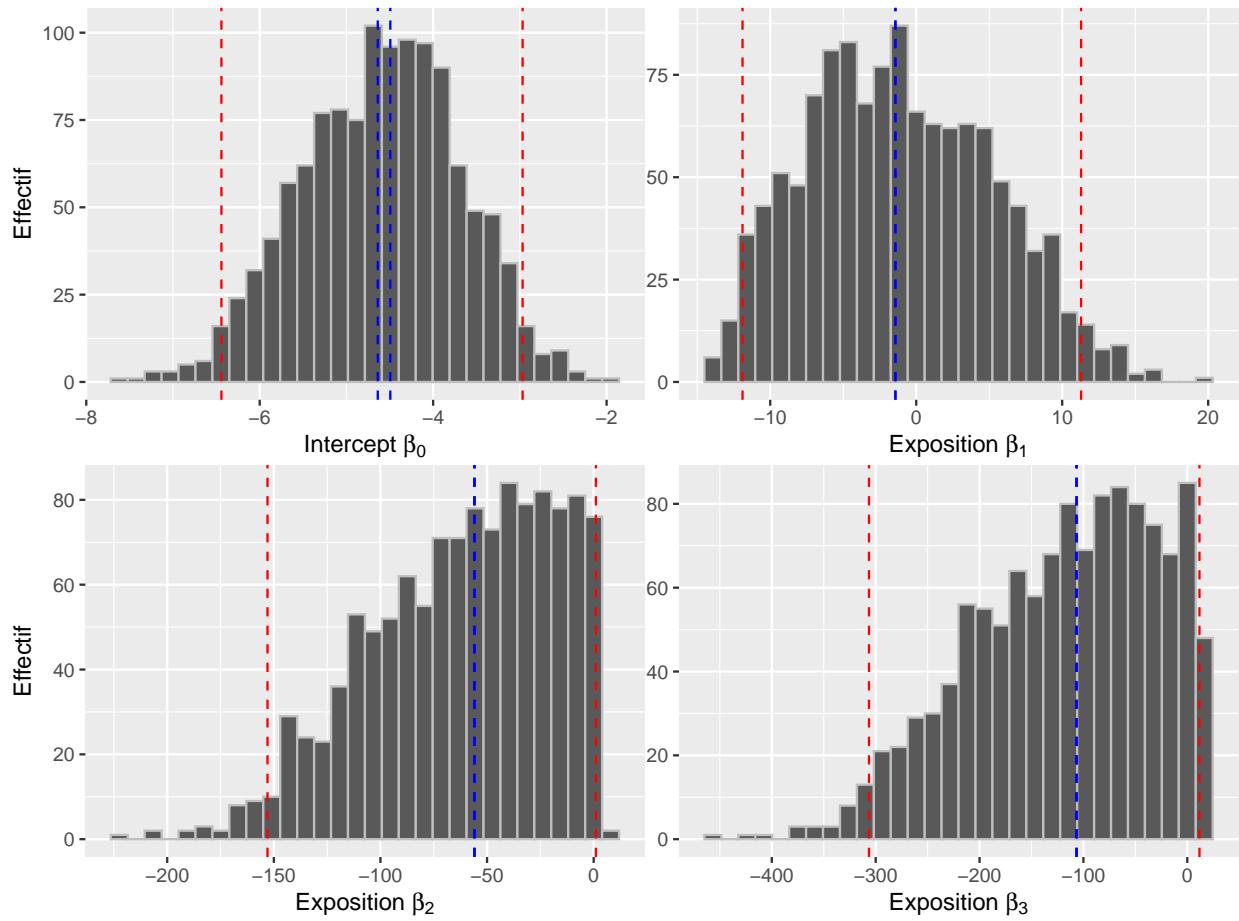


Figure 8: Intervalles de confiance dans le cas de normales univariées

2.2.2 Variation des coefficients selon une loi normale multivariée

A expliquer.

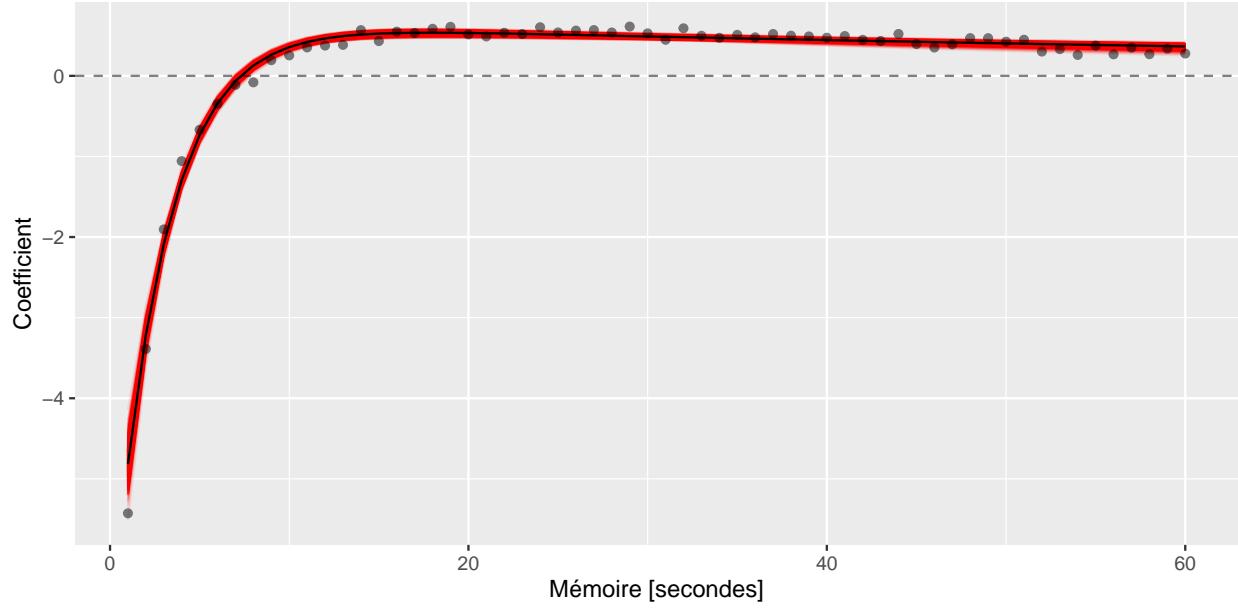


Figure 9: Variation des coefficients autorégressifs selon une loi normale multivariée

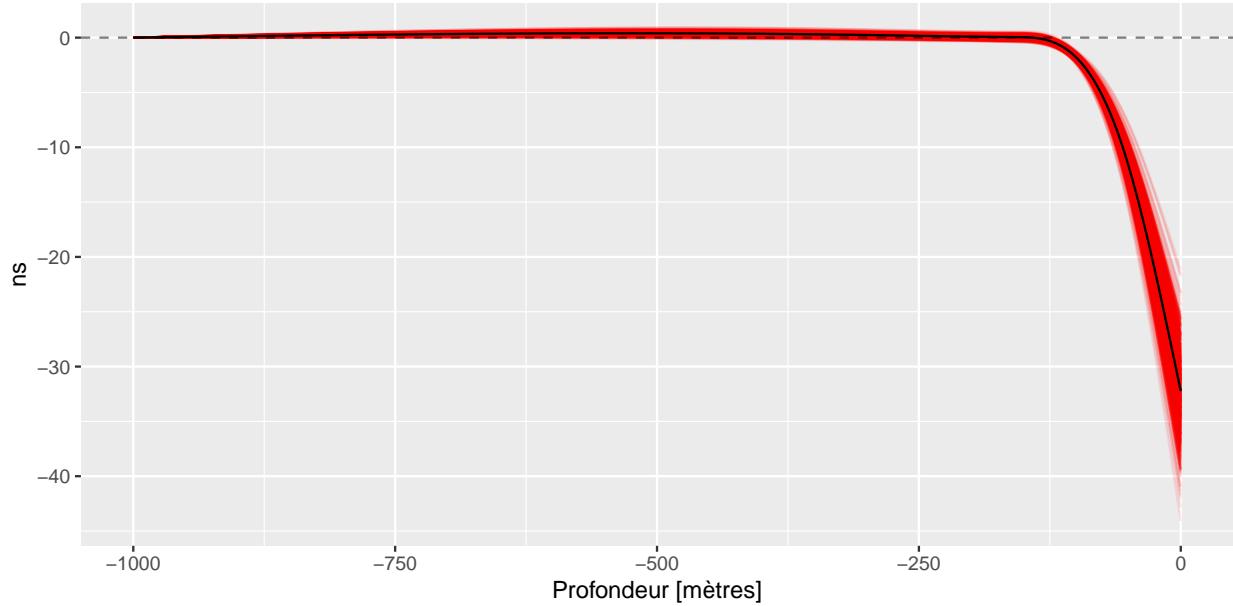


Figure 10: Variation des coefficients de profondeur selon une loi normale multivariée

Table 6: Intervalles de confiance dans le cas de normales multivariées

	MC - lower	MC - upper	SE - lower	SE - upper
(Intercept)	-4.887771	-4.273240	-4.631902	-4.497720
ns(X, knots = quantile(data\$X[data\$X > 0], 1:2/3))1	-1.971359	-0.480328	-1.237372	-1.228926
ns(X, knots = quantile(data\$X[data\$X > 0], 1:2/3))2	-64.614962	-52.023774	-58.309579	-58.096519
ns(X, knots = quantile(data\$X[data\$X > 0], 1:2/3))3	-124.492842	-98.741635	-111.627518	-111.208774

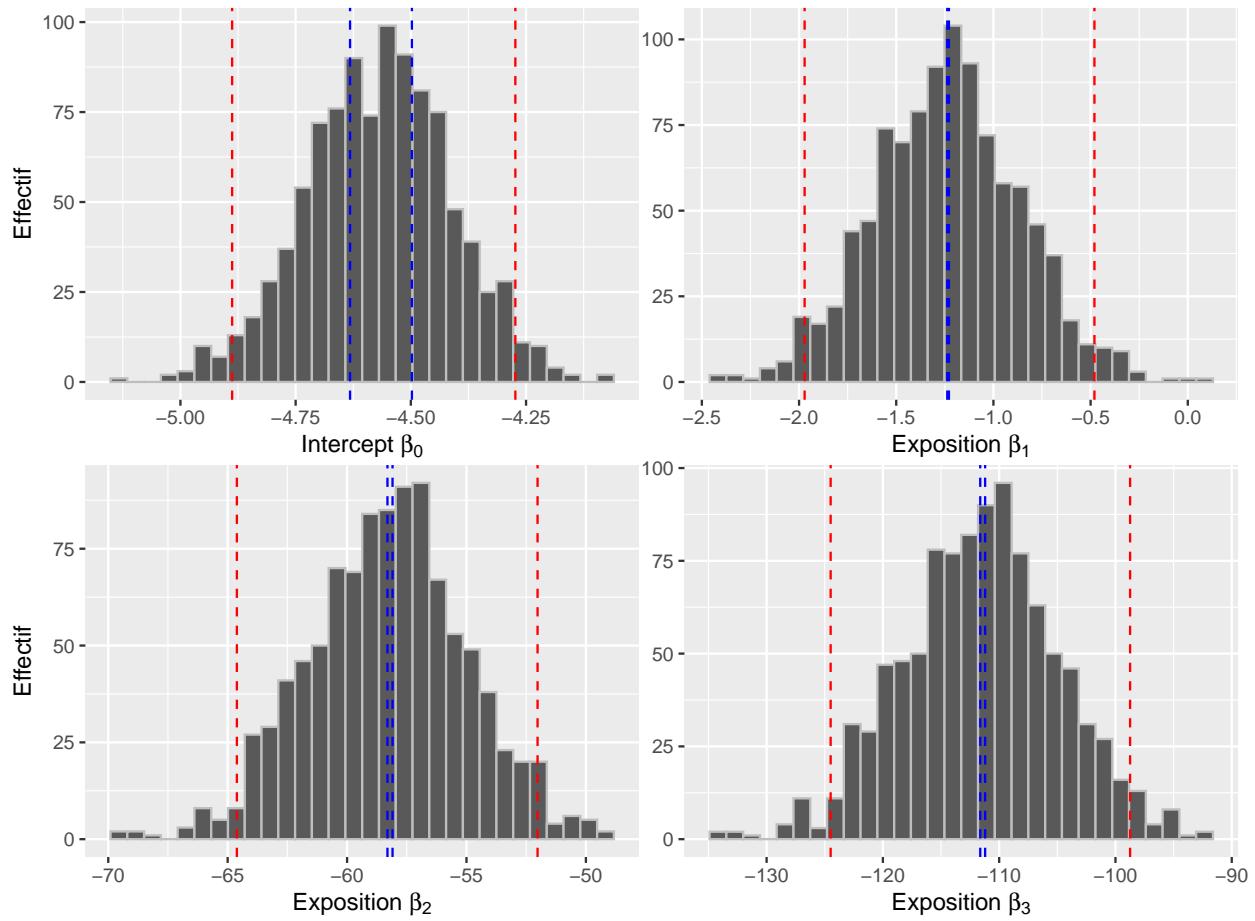


Figure 11: Intervalles de confiance dans le cas de normales multivariées

2.2.2.1 Covariance des coefficients autorégressifs et non des coefficients de la biexponentielle A expliquer.

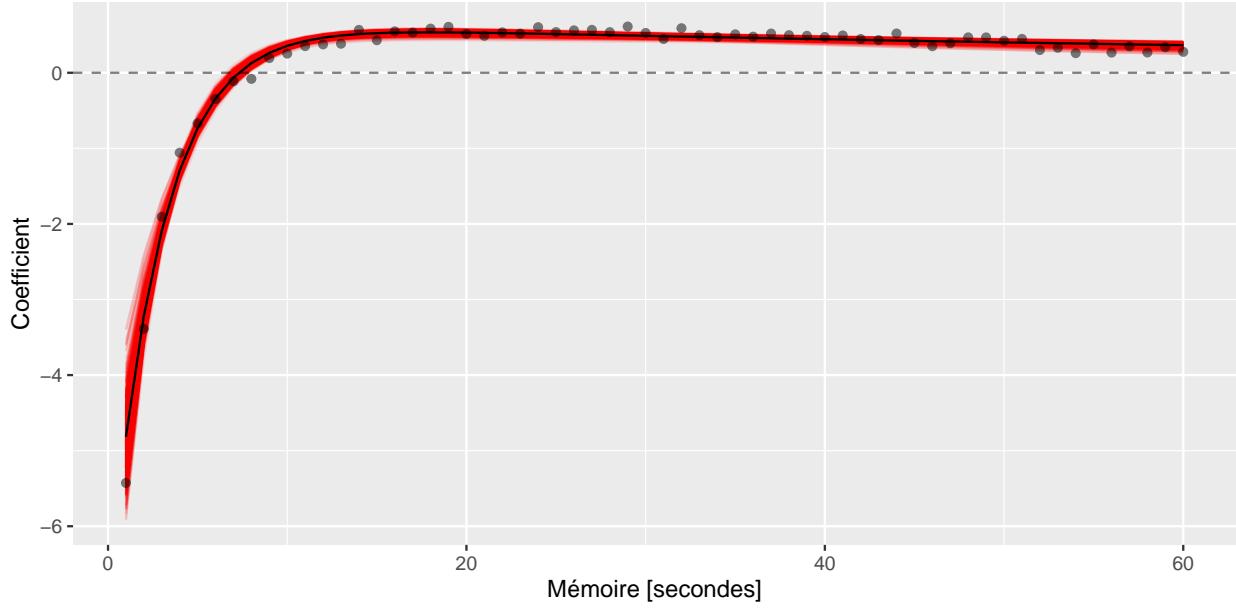


Figure 12: Variation des coefficients autorégressif selon une loi normale multivariée

Table 7: Intervalles de confiance dans le cas de normales multivariées

	MC - lower	MC - upper	SE - lower	SE - upper
(Intercept)	-4.865849	-4.2754368	-4.645331	-4.510474
ns(X, knots = quantile(data\$X[data\$X > 0], 1:2/3))1	-1.981092	-0.4261478	-1.122103	-1.113660
ns(X, knots = quantile(data\$X[data\$X > 0], 1:2/3))2	-65.140058	-51.9494486	-59.342551	-59.127834
ns(X, knots = quantile(data\$X[data\$X > 0], 1:2/3))3	-125.571605	-98.5452180	-113.695413	-113.273788

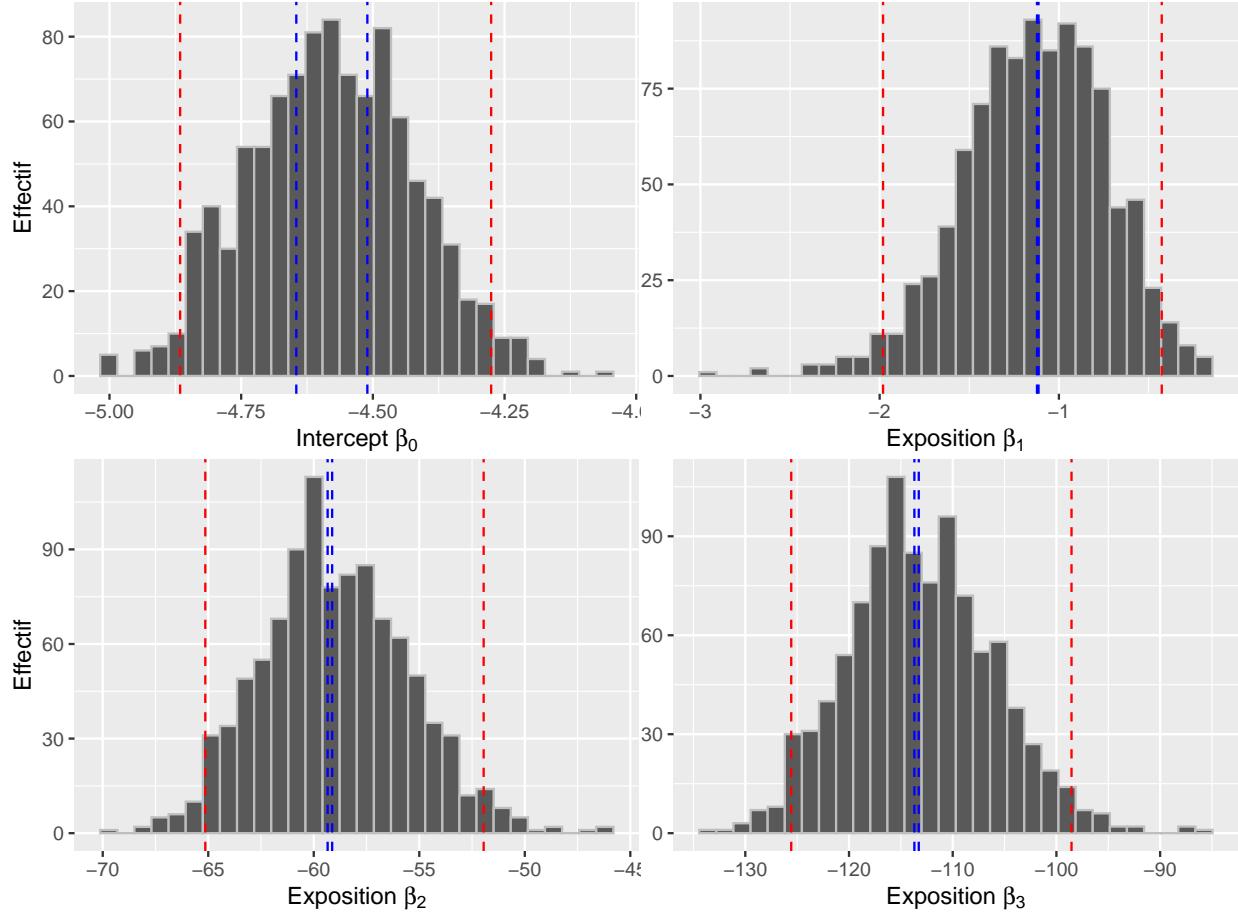


Figure 13: Intervalles de confiance dans le cas de normales multivariées

2.2.2.2 Sans passer par la régression biexponentielle A expliquer.

Table 8: Intervalles de confiance dans le cas de normales multivariées

	MC - lower	MC - upper	SE - lower	SE - upper
(Intercept)	-4.877449	-4.2607168	-4.654712	-4.512880
ns(X, knots = quantile(data\$X[data\$X > 0], 1:2/3))1	-1.782223	-0.5655889	-1.186667	-1.178238
ns(X, knots = quantile(data\$X[data\$X > 0], 1:2/3))2	-64.410381	-52.0926295	-58.180603	-57.969085
ns(X, knots = quantile(data\$X[data\$X > 0], 1:2/3))3	-124.003446	-99.0561959	-111.400643	-110.983884

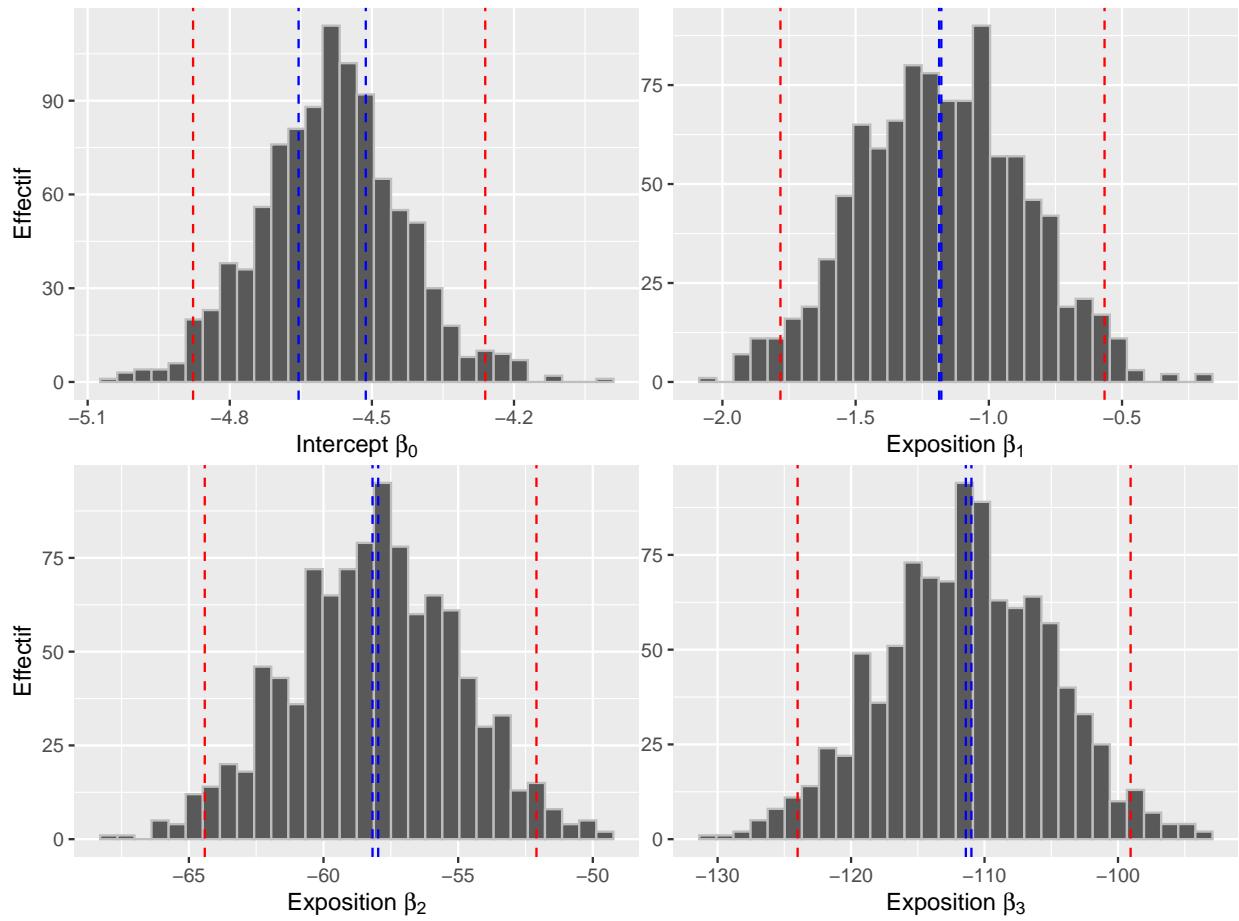


Figure 14: Intervalles de confiance dans le cas de normales multivariées