

Effet des perturbations humaines sur l'alimentation des narvals

Sommaire

1 Modélisation de l'effet de l'exposition sur le taux d'émission	2
1.1 Taux d'émission de buzz sans exposition	2
1.2 Effet de l'exposition	3
1.3 Construction d'intervalles de confiance	3
2 Résultats	4
2.1 Note sur le temps d'ajustement des modèles	4
2.2 Taux d'émission de buzz sans exposition	4
2.2.1 Recherche de la mémoire optimale	4
2.2.2 Régression double bi-exponentielle sur les coefficients autorégressifs	5
2.2.3 Coefficients de profondeur	6
2.3 Effet de l'exposition	6
2.4 Intervalles de confiance	7
2.4.1 Coefficients d'exposition	7
2.4.2 Pourcentage du taux normal d'émission de buzz	14

Les narvals sont des baleines vivant toute l'année au Groenland. Le réchauffement climatique favorise le recul des glaces sur le territoire groenlandais et ses côtes. Cela ouvre la porte au développement d'activités humaines au Groenland, et notamment les activités minières. Les biologistes se questionnent sur les effets potentiels engendrés sur les comportements des narvals.

Afin d'anticiper ces possibles changements de comportements une étude a été conduite pendant plusieurs mois en 2018. Dans ce cadre, 8 narvals ont été équipées de capteurs permettant d'enregistrer leur profondeur de plongée, leur localisation et les sons qu'elles émettent. Les baleines ont été laissées libres de toutes perturbations pendant plusieurs jours avant d'y être exposées. Les perturbations ont pris la forme de coups de fusil tirés dans l'eau depuis un bateau afin d'imiter les ondes émises par des activités minières.

Lorsqu'elles se nourrissent, les narvals émettent des sons spécifiques appelés "buzz". À partir des sons collectés il est donc possible de déterminer quand ces baleines sont en train de manger. La distance séparant les baleines du bateau émettant une perturbation peut-elle être calculée grâce aux puces GPS placées sur les narvals. Ainsi, nous pouvons tenter de modéliser l'effet de l'exposition des perturbations humaines sur la capacité des narvals à se nourrir.

1 Modélisation de l'effet de l'exposition sur le taux d'émission

1.1 Taux d'émission de buzz sans exposition

Etant donné que nous voulons modéliser un taux, la régression de Poisson est l'outil tout indiqué. On suppose que le taux d'émission de buzz Y , dépendant d'un ensemble de variables X , suit une loi de Poisson de paramètre λ , et on a que $\mathbb{E}(Y|X) = \lambda$. La régression de Poisson permet de modéliser le *log* de λ par une combinaison linéaire des variables X :

$$\log(\lambda) = \beta_0 + X\beta$$

Les variables X utilisées dans le modèle sont détaillées dans la suite de cette section. Il n'existe pas d'expression explicite de l'estimateur du maximum de vraisemblance des paramètres β , cependant maximiser cette vraisemblance est un problème d'optimisation convexe qui peut être résolu numériquement.

Les données que nous utilisons correspondent à plusieurs individus, nous avons donc plusieurs observations par individus et celles-ci ne sont pas indépendantes. Pour palier ce défaut de modélisation et tenir compte de la spécifité des individus, nous avons ajouté des effets aléatoires b_i au modèle :

$$\log(\lambda_i) = \beta_0 + b_i + X_i\beta$$

où i dénote les observations de l'individu i et $b_i \sim \mathcal{N}(0, \sigma_i^2)$, σ_i^2 étant la variance intra-individu.

Pour se nourrir, les narvals doivent plonger profondément (plusieurs centaines de mètres), alors que le reste du temps elles restent "proche" (quelques dizaines de mètres) 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 spline cubique naturelle ayant pour noeuds les quantiles 1/3 et 2/3. Nous pouvons expliciter un peu plus l'expression précédente du modèle :

$$\log(\lambda_i) = \beta_0 + b_i + \text{spline}(\text{Depth}_i)\beta_D$$

Passer par l'estimateur du maximum de vraisemblance nécessite que les observations Y_{ij} soient indépendantes. Or, l'émission d'un buzz à un instant t est corrélé à l'émission ou non de buzz aux instants précédents ; cet effet mémoire doit donc être intégré au modèle pour rendre les Y_{ij} indépendants. Pour cela nous introduisons K variables binaires d'auto-régression codant l'émission d'un buzz aux instants $t-k$, $k \in \{1, \dots, K\}$. Le modèle résultant s'écrit alors :

$$\log(\lambda_i(t)) = \beta_0 + b_i + \text{spline}(\text{Depth}_i(t))\beta_D + \sum_{k=1}^K \alpha_k Y_i(t-k)$$

Cette approche demande de fixer une mémoire maximale, et ainsi la valeur de K . Pour choisir la mémoire maximale optimale, nous avons fait varier K et utilisé le BIC comme mesure de la qualité des différents modèles correspondants ; nous choisissons la mémoire maximale K_{opt} du modèle minimisant ce critère. Pour éviter de parcourir tout l'ensemble $\{K_{min}, \dots, K_{max}\}$, nous avons utilisé une démarche permettant de restreindre le pas de recherche au fur et à mesure que l'on s'approche de K_{opt} :

1. $from.k = K_{min}$; $to.k = K_{max}$
2. Tant que $(to.k - from.k > 2)$:
 1. Pour $K_i = from.k + (i - 1) * \lfloor \frac{to.k - from.k}{M-1} \rfloor$, $i \in \{1, \dots, M\}$:
 1. ajustement d'un modèle avec K_i éléments mémoire
 2. calcul du BIC
 2. $i_{opt} = argmin BIC_i$
 3. $K_{opt} = K_{i_{opt}}$
 4. $from.k = K_{i_{opt}-1}$; $to.k = K_{i_{opt}+1}$

Le nombre de composants auto-régressifs pouvant être grand, nous avons utilisé une régression bi-exponentielle double permettant de lier le décalage mémoire et le coefficient associé :

$$BiExp(lag) = A_1 e^{-e^{lrc_1} lag} + A_2 e^{-e^{lrc_2} lag}$$

Cela permet de réduire le nombre de coefficients de K_{opt} à 4, ce qui est doublement bénéfique : le temps d'ajustement des modèles est grandement réduit et lors de la construction des intervalles de confiance de nos coefficients, l'accumulation des variances est limitée.

1.2 Effet de l'exposition

Le niveau d'exposition aux perturbations est représenté par l'inverse de la distance séparant la baleine du bateau quand un coup de feu est tiré. De même que pour la profondeur, la non-linéarité de la relation entre le niveau d'exposition et le taux d'émission de buzz est représentée par l'utilisation d'une spline cubique naturelle dont les noeuds sont les quantiles 1/3 et 2/3 des niveaux d'exposition.

Afin d'observer l'effet de l'exposition aux perturbations par rapport à l'émission de buzz sans perturbation, nous ajoutons un terme d'"offset". Cet offset est calculé à partir des coefficients $\hat{\beta}_D$, \hat{A}_1 , \hat{lrc}_1 , \hat{A}_2 , \hat{lrc}_2 estimés sans perturbation :

$$offset_i(t) = spline(Depth_i(t))\hat{\beta}_D + \sum_{k=1}^{K_{opt}} (\hat{A}_1 e^{-e^{\hat{lrc}_1} k} + \hat{A}_2 e^{-e^{\hat{lrc}_2} k}) Y_i(t - k)$$

Le modèle incluant l'exposition s'écrit donc :

$$\log(\lambda_i(t)) = offset_i(t) + \beta_0 + b_i + spline(Expo_i(t))\beta_E$$

Et les paramètres estimés $\hat{\beta}_E$ permettront d'interpréter l'effet de l'exposition par rapport à des conditions de stress normales.

1.3 Construction d'intervalles de confiance

Pour se rassurer quant à l'interprétabilité des résultats obtenus il est important de construire les intervalles de confiance des estimations des coefficients associés à l'exposition aux perturbations. L'approche classique de calcul des intervalles de confiance se basant sur la seule variance estimée des coefficients d'exposition donnerait ici des résultats incorrects, la variance des coefficients de profondeur et d'auto-régression ne serait

alors pas prise en compte car tuée par l'utilisation de l'offset. Nous avons donc utilisée une approche Monte-Carlo lors de laquelle nous répétons $M = 1000$ fois l'ajustement du modèle incluant l'exposition en tirant aléatoirement les coefficients $\hat{\beta}_D$, \hat{A}_1 , \hat{lrc}_1 , \hat{A}_2 , \hat{lrc}_2 selon leur moyenne et leur variance. Les 1000 ajustements obtenus nous fournissent ainsi un échantillon pour chacun des coefficients d'exposition et nous construisons leur intervalle de confiance via les quantiles empiriques de leur distribution.

2 Résultats

2.1 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 1 cette augmentation est linéaire pour les modèles classiques, alors que pour les modèles mixtes celle-ci est quadratique.

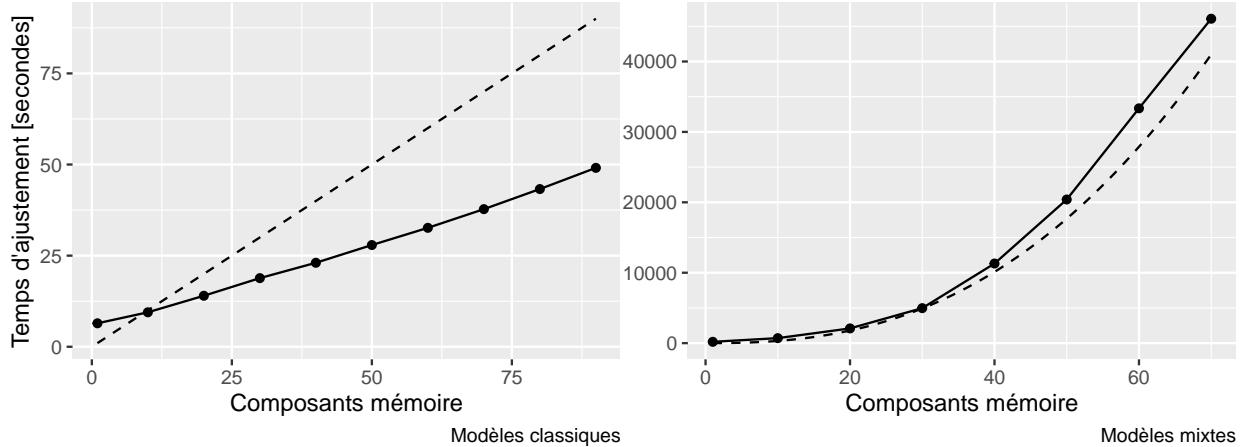


Figure 1: Temps d'ajustement (en secondes) des modèles en fonction de la mémoire maximum.

2.2 Taux d'émission de buzz sans exposition

2.2.1 Recherche de la mémoire optimale

Le temps d'ajustement des modèles mixtes étant nettement plus important que ceux des modèles sans effets aléatoires, nous avons dans un premier temps exclu ces effets du modèle sans exposition pour estimer la mémoire optimale. Nous avons choisi pour la recherche $K_{min} = 1$, $K_{max} = 300$, $M = 10$. Nous obtenons une mémoire optimale de 60 secondes. La figure 2 permet de voir que cet optimal semble bien correspondre à un minimum global.

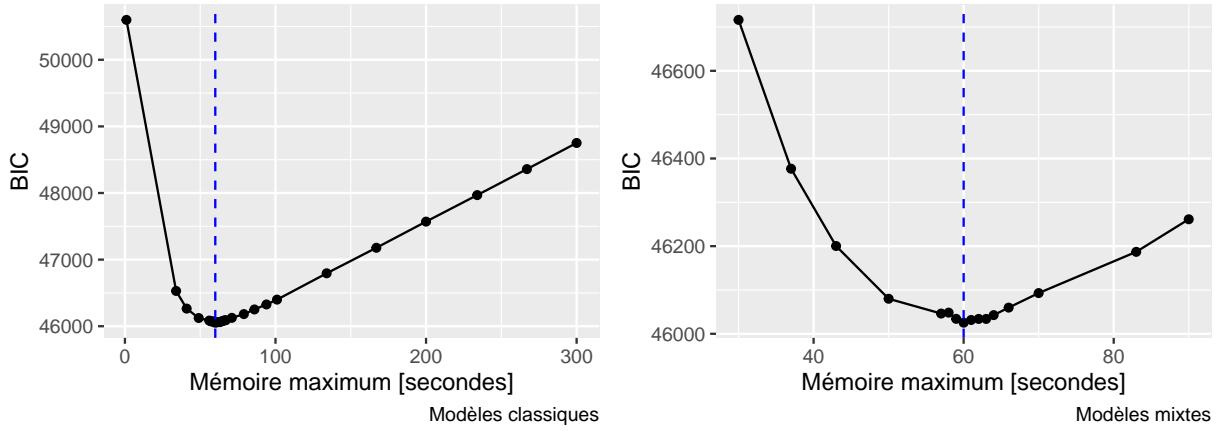


Figure 2: BIC en fonction de la mémoire maximum.

Cela nous a permis de restreindre tout de suite l'ensemble de recherche initial à $\{30, \dots, 90\}$ quand nous avons considéré le modèle incluant les effets aléatoires. De même que précédemment, nous obtenons une mémoire optimale de 60 secondes et la figure ?? nous indique qu'il s'agit bien d'un optimum global.

2.2.2 Régression double bi-exponentielle sur les coefficients autorégressifs

La figure 3 permet de comparer la régression double bi-exponentielle et les composantes de la mémoire ajustées pour un décalage maximum de 60. Nous pouvons constater que l'ajustement par la double bi-exponentielle est très fidèle aux 60 coefficients initiaux.

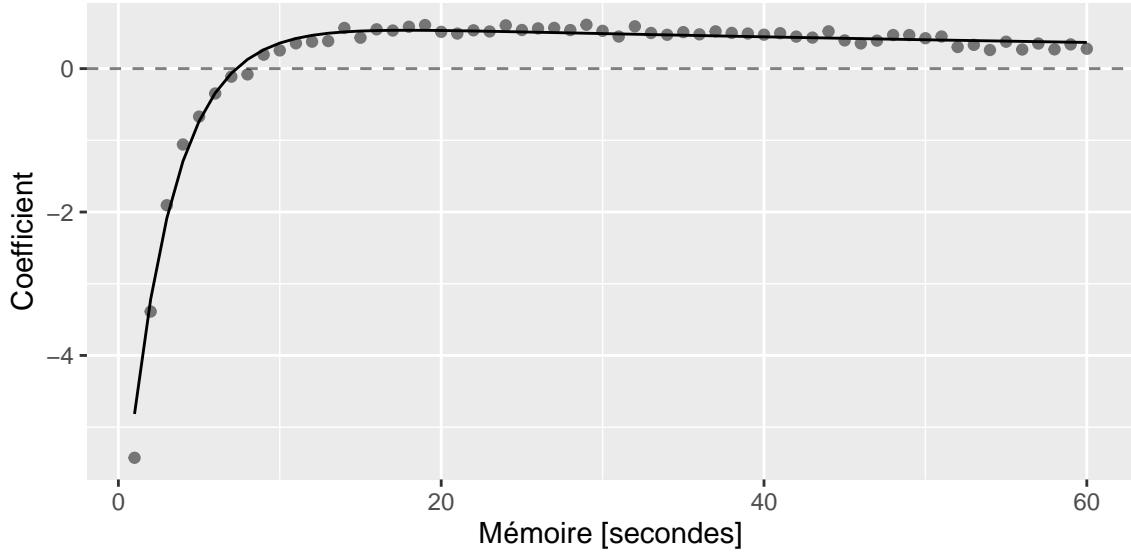


Figure 3: Régression bi-exponentielle des coefficients auto-régressifs.

Les 4 coefficients obtenus suite à la régression double bi-exponentielle sont présentés dans la table 1. Nous pouvons noter que leurs erreurs standard sont faibles, ce qui est cohérent avec la validation visuelle de la régression.

Table 1: Coefficients autorégressifs obtenus par régression bi-exponentielle

	moyenne	erreur standard	p-valeur
A_1	-7.7413301	0.3287556	0
lrc_1	-1.0545001	0.0407916	0
A_2	0.6544679	0.0427976	0
lrc_2	-4.6329847	0.1880983	0

2.2.3 Coefficients de profondeur

Les coefficients de profondeur du modèle linéaire mixte sans exposition sont proposés table 2. Interpréter les coefficients associés à des splines étant difficile, nous nous contenterons de relever que seul les pvalues de β_{D_2} et β_{D_3} sont significatives au seuil $\alpha = 5\%$ et que l'erreur standard de β_{D_1} semble élevée relativement à sa moyenne.

Table 2: Coefficients de profondeur.

	moyenne	erreur standard	p-valeur
β_{D_1}	-0.1254979	0.2069123	0.5441647
β_{D_2}	-23.4762851	2.4222863	0.0000000
β_{D_3}	-32.6394721	3.3383763	0.0000000

2.3 Effet de l'exposition

La table 3 expose les coefficients associés à l'exposition du modèle mixte. De même que pour les coefficients de profondeur, il est compliqué d'interpréter leur valeur. Aussi, nous préférons nous référer à la lecture de la figure 4 sur laquelle nous représentons les estimations du modèle pour le taux d'émission de buzz selon la distance avec le bateau. Il est clair que plus le bateau est loin, plus le taux d'émission est élevé.

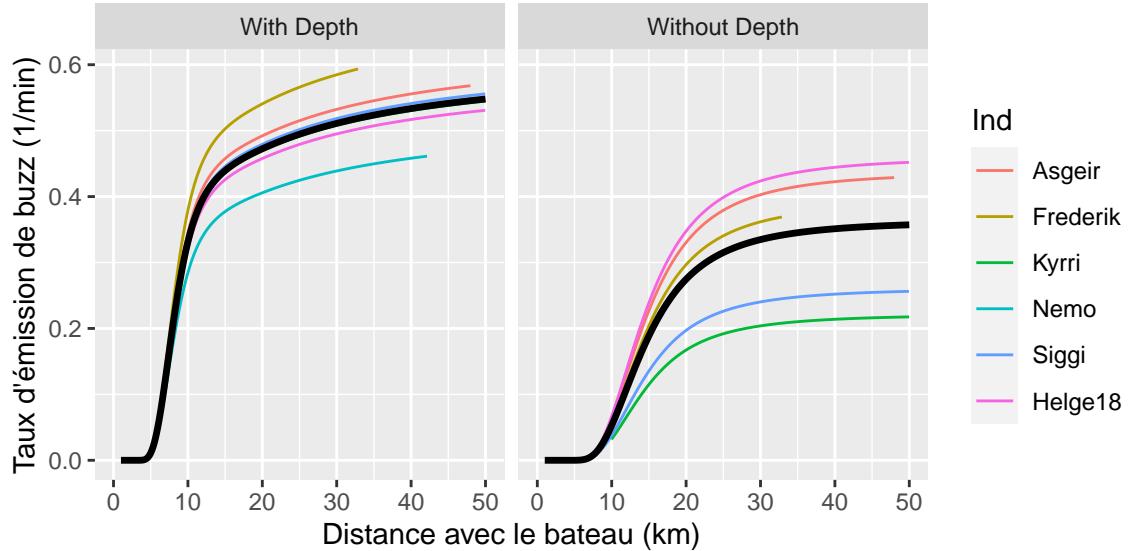


Figure 4: Evolution du taux d'émission de buzz selon la distance au bateau.

Table 3: Coefficients d'exposition.

	moyenne	erreur standard	p-valeur
β_{E_1}	-1.145812	0.0021544	0
β_{E_2}	-58.872353	0.0543014	0
β_{E_3}	-112.790469	0.1069174	0

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.

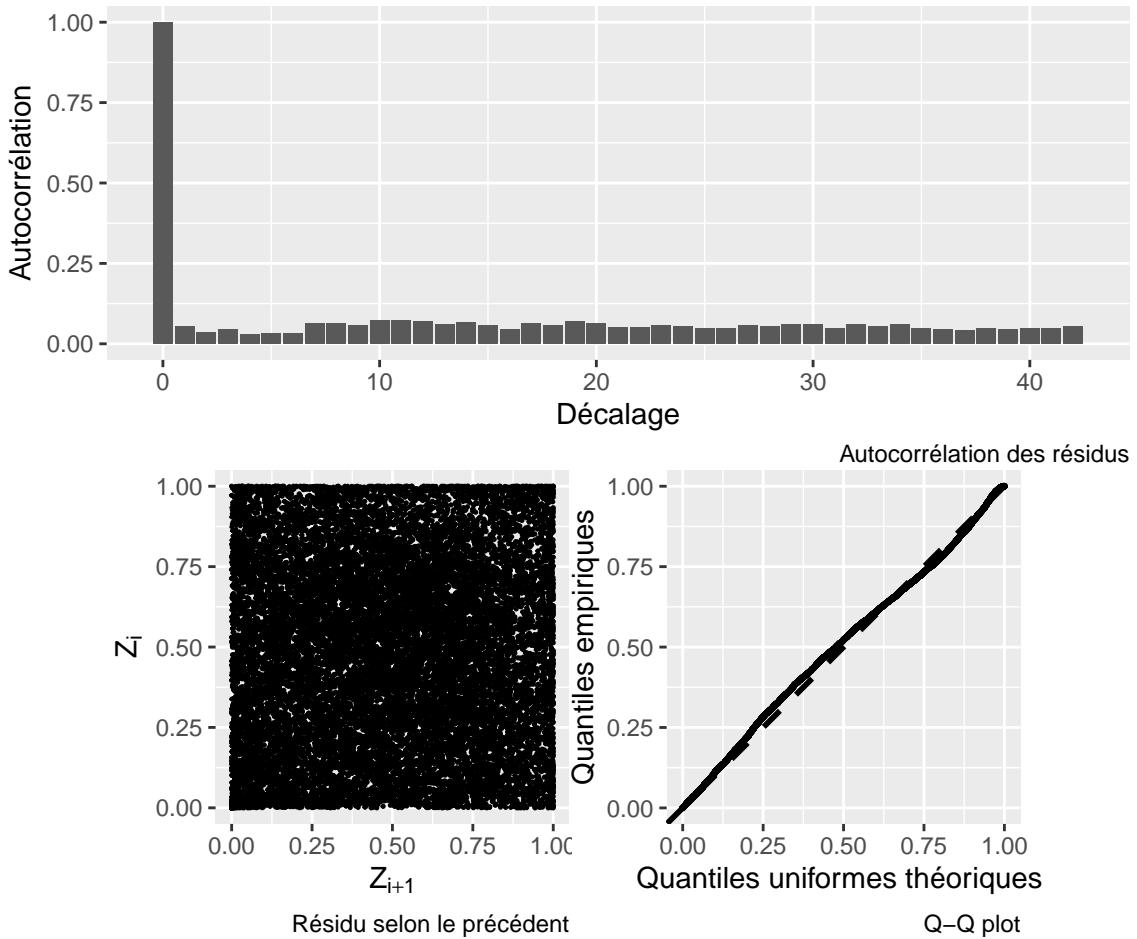


Figure 5: Validation graphique du modèle.

2.4 Intervalles de confiance

2.4.1 Coefficients d'exposition

2.4.1.1 Utilisation de lois normales univariées Dans un premier temps, nous avons considéré que les coefficients autorégressifs et de profondeur suivaient chacun une loi normale centrée sur leur estimation moyenne et avec une variance égale au carré de leur erreur standard. Nous avons donc 4 lois normales

univariées pour les coefficients de la régression double bi-exponentielle et 4 autres pour la spline sur la profondeur.

Sur la figure 6 nous avons représenté les courbes des fonctions double bi-exponentielles ainsi générées. Nous pouvons voir que leurs allures semblent toujours suivre celle de la régression initiale.

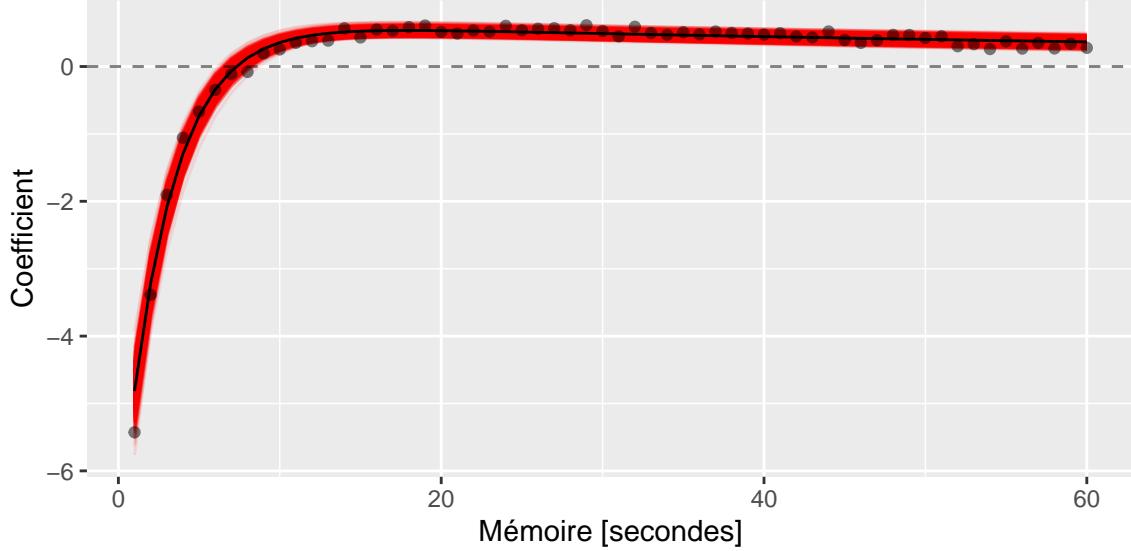


Figure 6: Variation des coefficients autorégressif selon des lois normales univariées.

Nous avons fait de même avec les coefficients de profondeur, et nous pouvons remarquer sur la figure 7 que cette fois-ci l'allure de certaines courbes s'éloignent sensiblement de l'interpolation moyenne.

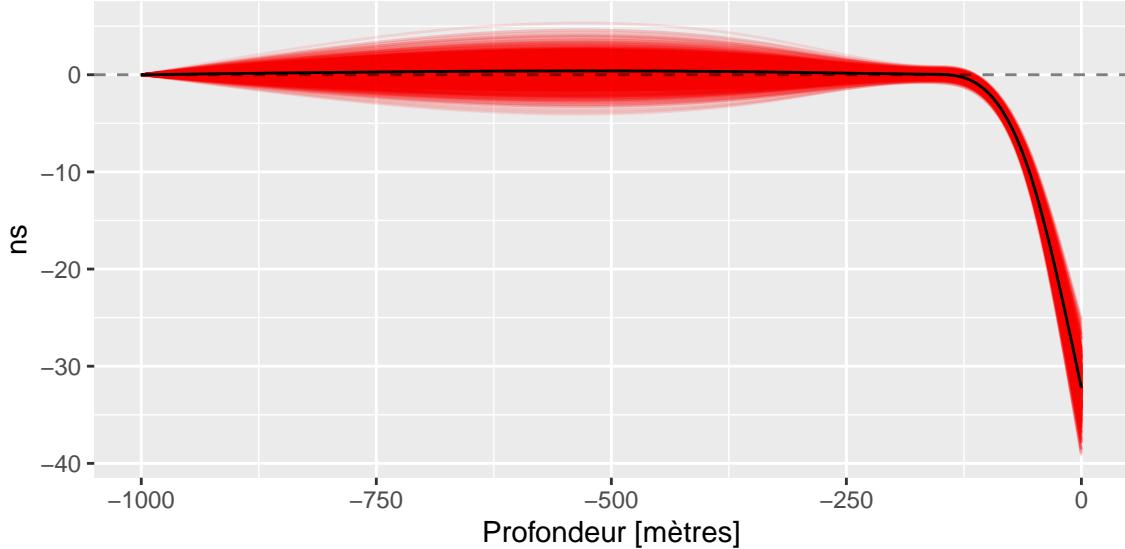


Figure 7: Variation des coefficients de profondeur selon des lois normales univariées.

Les intervalles de confiance calculés avec la procédure Monte Carlo sont donnés sur la table 4. Nous avons également affiché les valeurs médianes des intervalles de confiance calculés à chaque répétition sur la base

de la variation des coefficients d'exposition uniquement. Il est flagrant que les intervalles Monte Carlo sont bien plus larges et ne permettent en aucun cas de conclure sur un effet de l'exposition sur le taux d'émission de buzz.

Table 4: Intervalles de confiance dans le cas de normales univariées.

	Monte Carlo		Erreur standard	
	inf	sup	inf	sup
β_0	-6.441511	-2.968321	-4.638673	-4.494102
β_{E_1}	-11.898716	11.294079	-1.430436	-1.422075
β_{E_2}	-152.959472	1.047762	-56.014702	-55.803478
β_{E_3}	-306.428946	11.706679	-106.846662	-106.430830

La figure 8 donne une représentation graphique de ces intervalles (en rouge les “MC” et en bleu les “SE”), ainsi que des distributions des coefficients.

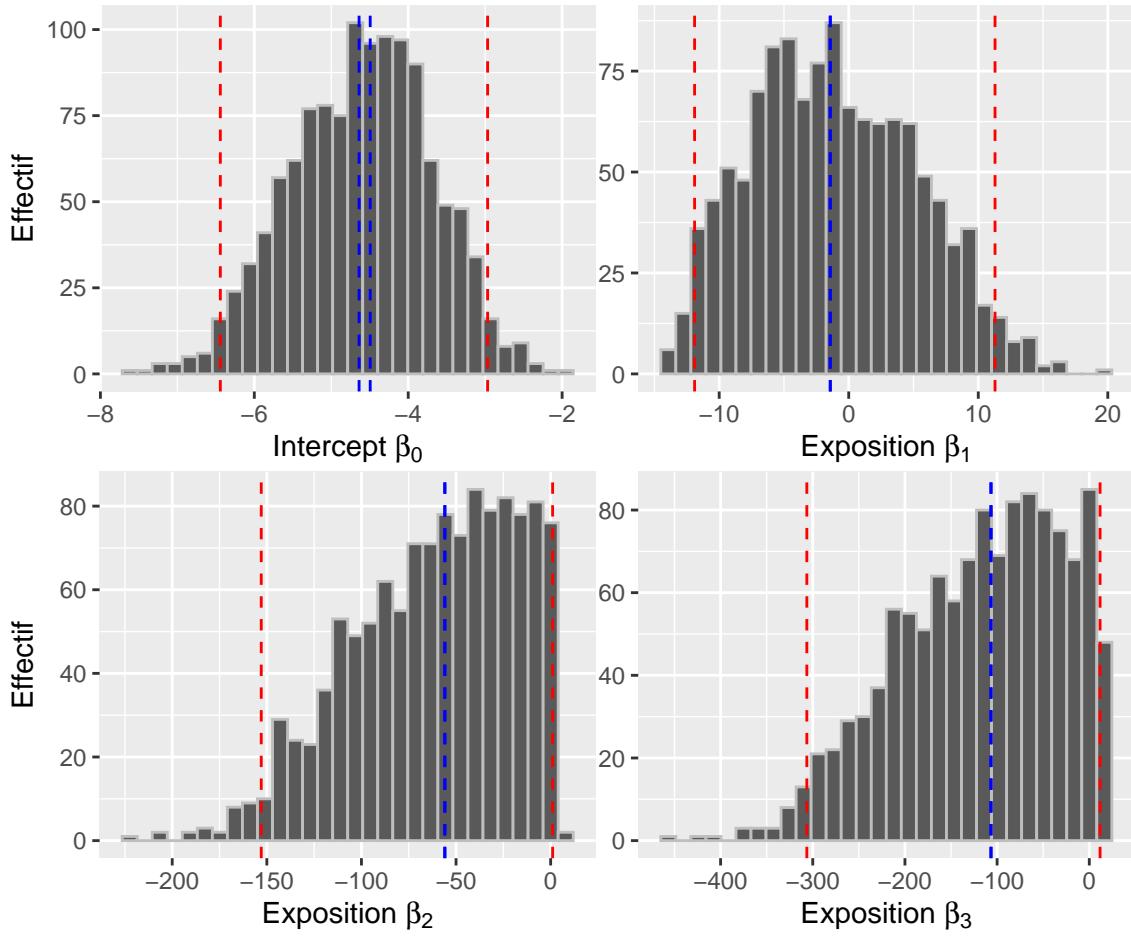


Figure 8: Intervalles de confiance dans le cas de normales univariées.

2.4.1.2 Utilisation d'une loi normale multivariée L'une des explications aux très larges intervalles de confiance observés dans la section précédente pourrait être que nous avons accumulé les variances des coefficients sans tenir compte des probables covariances existantes entre les coefficients. Afin de corriger cela

nous avons répété l'approche décrite précédemment, mais en tirant les coefficients dans des lois normales multivariées (une pour les coefficients de la régression double bi-exponentielle et une pour les coefficients de la spline) toujours centrées sur les estimations moyennes, mais ayant des matrices de variance-covariance non diagonales. Dorénavant nous avons donc une loi normale multivariée pour les coefficients de la régression double bi-exponentielle et une autre pour la profondeur, les deux étant de dimension 4.

La figure 9 montre que la régression biexponentielle initiale est encore plus fidèlement suivie qu'auparavant.

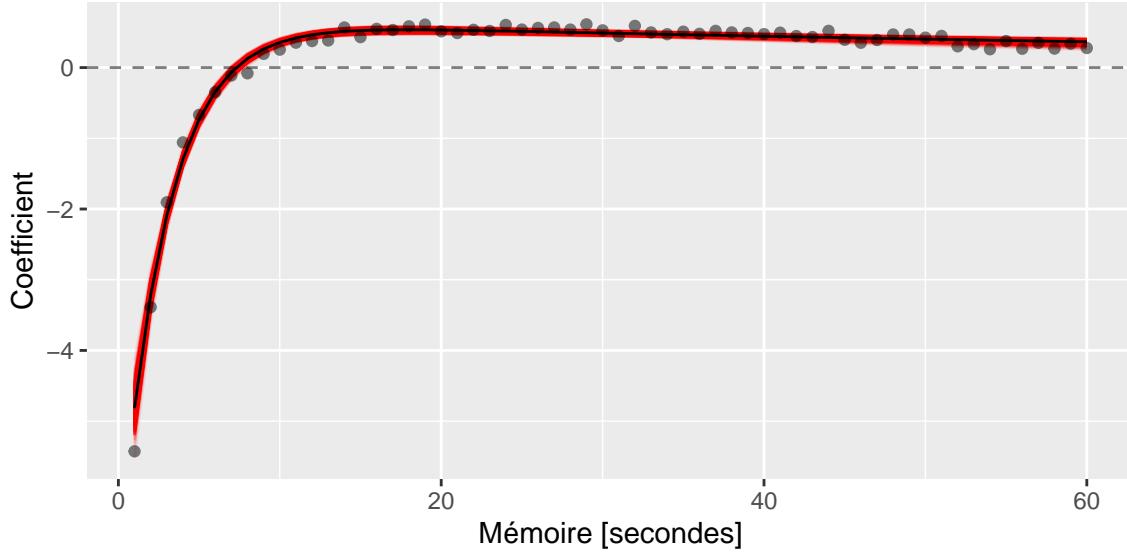


Figure 9: Variation des coefficients autorégressifs selon une loi normale multivariée.

Et surtout, comme nous pouvons le voir sur la figure 10, il en va de même pour la profondeur, alors que précédemment les tirages donnaient des courbes fortement éloignées de celle attendue.

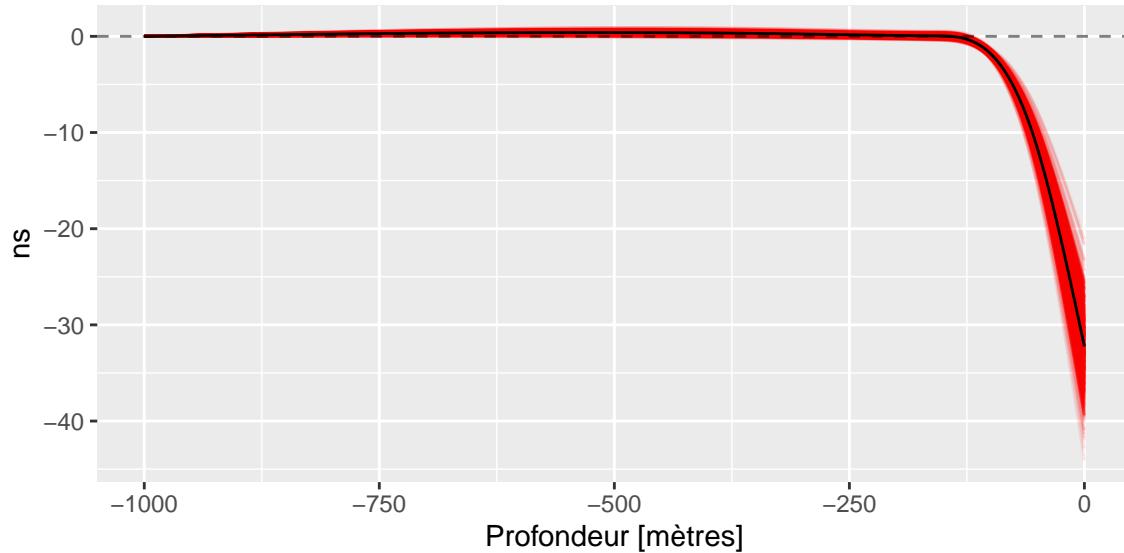


Figure 10: Variation des coefficients de profondeur selon une loi normale multivariée.

Sur la table 5 nous pouvons constater que les intervalles de confiance estimés en utilisant des lois normales

multivariées sont nettement plus petits et peuvent conduire à conclure sur un effet de l'exposition sur le taux d'émission de buzz.

Table 5: Intervalles de confiance dans le cas de normales multivariées.

	Monte Carlo		Erreur standard	
	inf	sup	inf	sup
β_0	-4.887771	-4.273240	-4.631902	-4.497720
β_{E_1}	-1.971359	-0.480328	-1.237372	-1.228926
β_{E_2}	-64.614962	-52.023774	-58.309579	-58.096519
β_{E_3}	-124.492842	-98.741635	-111.627518	-111.208774

Nous pouvons également voir sur la figure 11 que les distributions semblent normales, contrairement à celles observées avec les lois univariées.

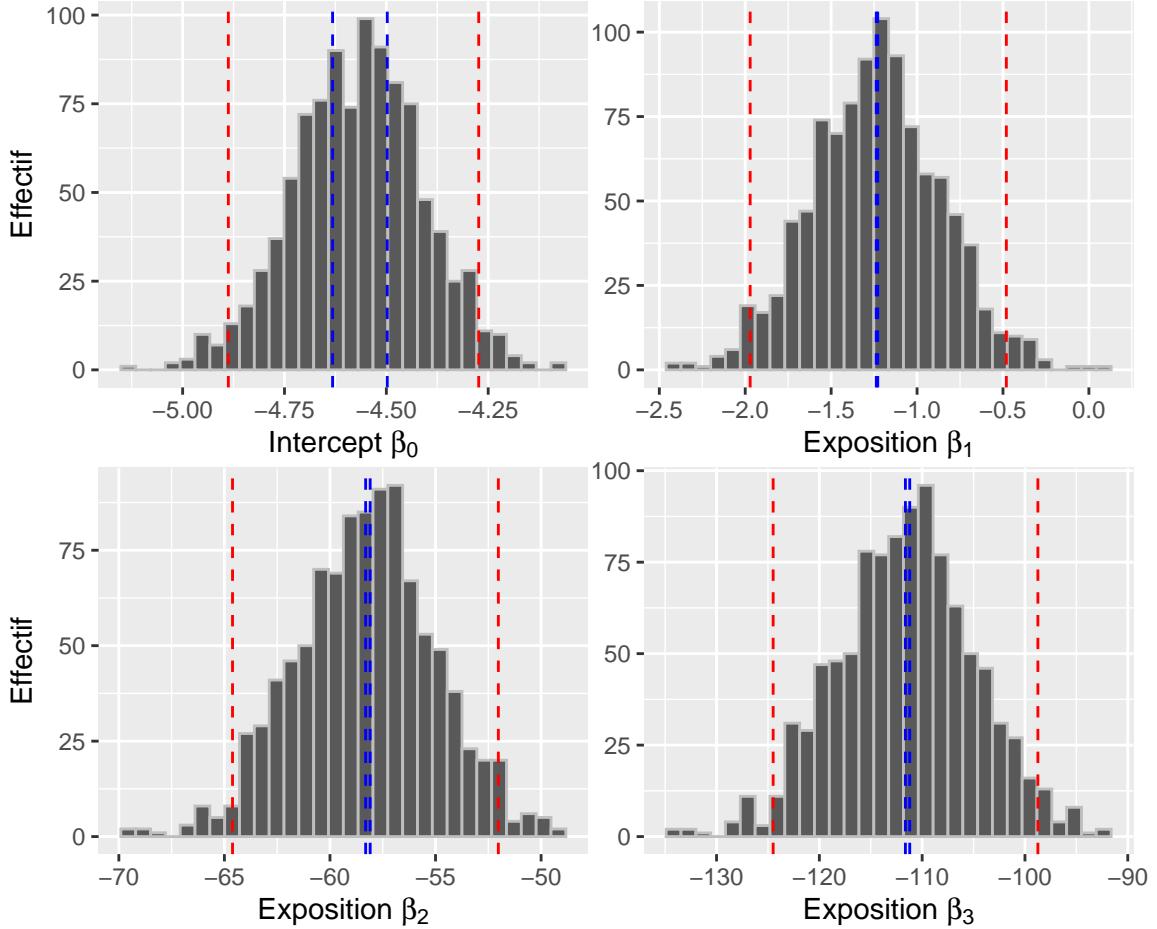


Figure 11: Intervalles de confiance dans le cas de normales multivariées.

2.4.1.3 Variance-covariance des coefficients autorégressifs Bien que plus intéressante, la procédure que nous avons mise en place se base sur la matrice de variance-covariance des coefficients de la régression double bi-exponentielle, ce qui signifie que nous ne captions pas directement la variabilité du phénomène autorégressif mais plutôt celle de la régression double bi-exponentielle. Pour palier cette approximation nous

avons tenté d'estimer la matrice de variance-covariance des coefficients autorégressifs, à nouveau par une procédure Monte Carlo. Cette fois-ci nous répétons 1000 fois le tirage des 60 coefficients de mémoire à partir d'une loi normale multivariée et à chaque fois nous ajustons une régression double bi-exponentielle sur les 60 coefficients tirés ; finalement nous pouvons obtenir les estimations moyennes des coefficients de la double bi-exponentielle, ainsi que la matrice de variance-covariance associée. Nous reprenons enfin la procédure Monte Carlo employée dans la section précédente avec le vecteur de moyennes et la matrice de variance-covariance obtenus.

Nous pouvons contrôler sur la figure 12 que les régressions double bi-exponentielles obtenues restent cohérentes.

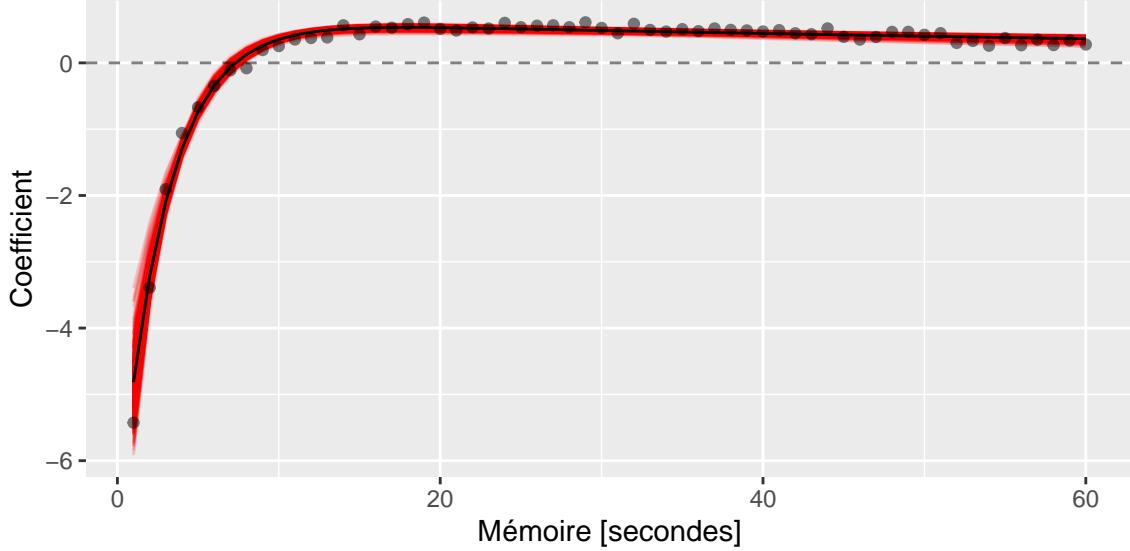


Figure 12: Variation des coefficients autorégressif selon une loi normale multivariée estimée par Monte Carlo.

La table 6 et la figure 13 permettent de vérifier que bien que les intervalles calculés ainsi sont légèrement plus grands que les précédents, ils permettent toujours d'interpréter les coefficients d'exposition ajustés par le modèle.

Table 6: Intervalles de confiance dans le cas d'une normale multivariée estimée par Monte Carlo.

	Monte Carlo		Erreur standard	
	inf	sup	inf	sup
β_0	-4.865849	-4.2754368	-4.645331	-4.510474
β_{E_1}	-1.981092	-0.4261478	-1.122103	-1.113660
β_{E_2}	-65.140058	-51.9494486	-59.342551	-59.127834
β_{E_3}	-125.571605	-98.5452180	-113.695413	-113.273788

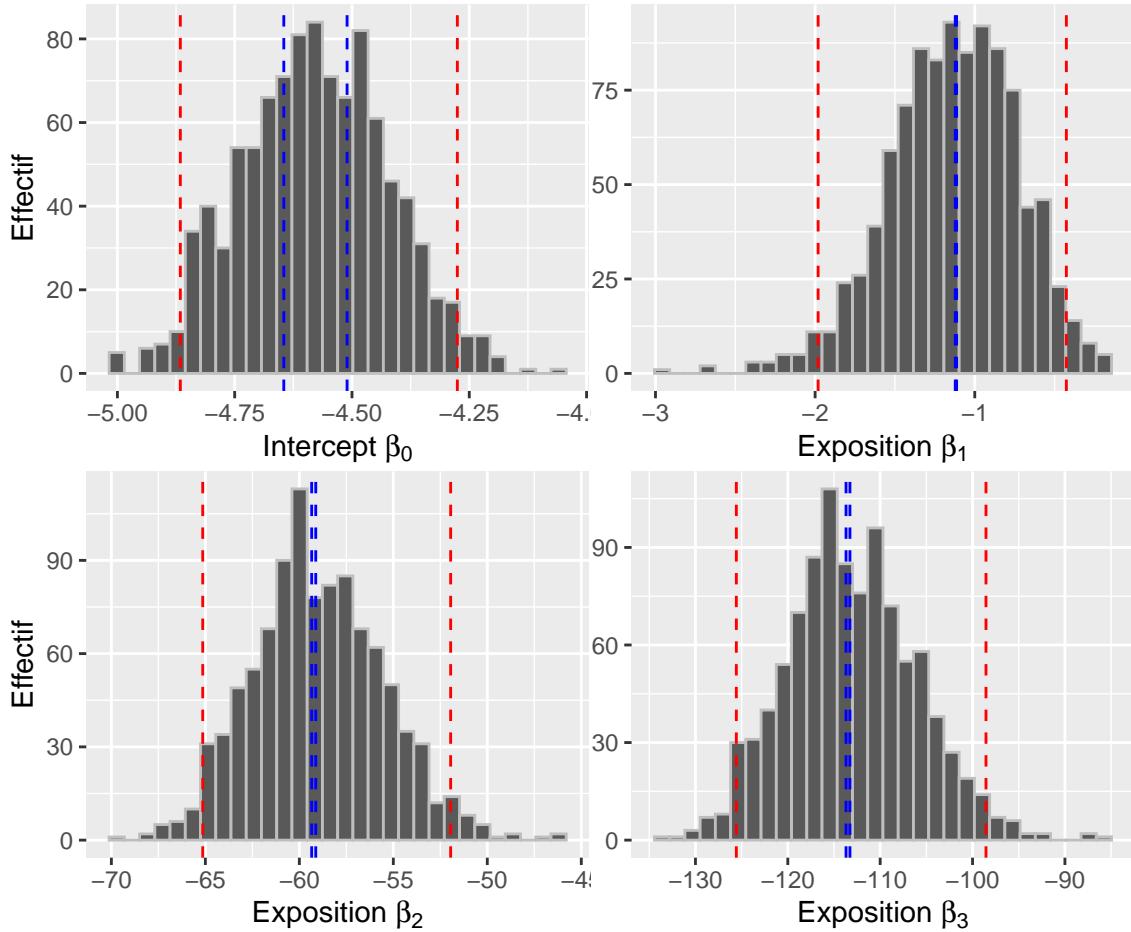


Figure 13: Intervalles de confiance dans le cas d'une normale multivariée estimée par Monte Carlo.

2.4.1.4 Sans passer par la régression biexponentielle L'intérêt de la régression biexponentielle était de :

1. réduire le temps d'ajustement des modèles linéaires,
2. éviter d'accumuler les variances des 60 coefficients de mémoire.

En fixant les coefficients autorégressifs, avoir 4 ou 60 coefficients pour la mémoire n'importe plus ; et en utilisant une loi normale multivariée, nous devrions également ne plus accumuler directement les variances des coefficients. Ainsi, nous avons tenté de nous passer de la régression double bi-exponentielle, et donc de tirer les 60 coefficients mémoire et les 4 coefficients de profondeur directement dans une seule loi normale multivariée.

Nous pouvons voir sur la table 7 et la figure 14 qu'avec cette approche plus directe, les intervalles de confiance sont quasiment identiques à ceux obtenus dans la section 2.4.1.2, et même plus petit pour l'intercept.

Table 7: Intervalles de confiance dans le cas d'une loi normale multivariée sans double bi-exponentielle.

	Monte Carlo		Erreur standard	
	inf	sup	inf	sup
β_0	-4.877449	-4.2607168	-4.654712	-4.512880
β_{E_1}	-1.782223	-0.5655889	-1.186667	-1.178238
β_{E_2}	-64.410381	-52.0926295	-58.180603	-57.969085
β_{E_3}	-124.003446	-99.0561959	-111.400643	-110.983884

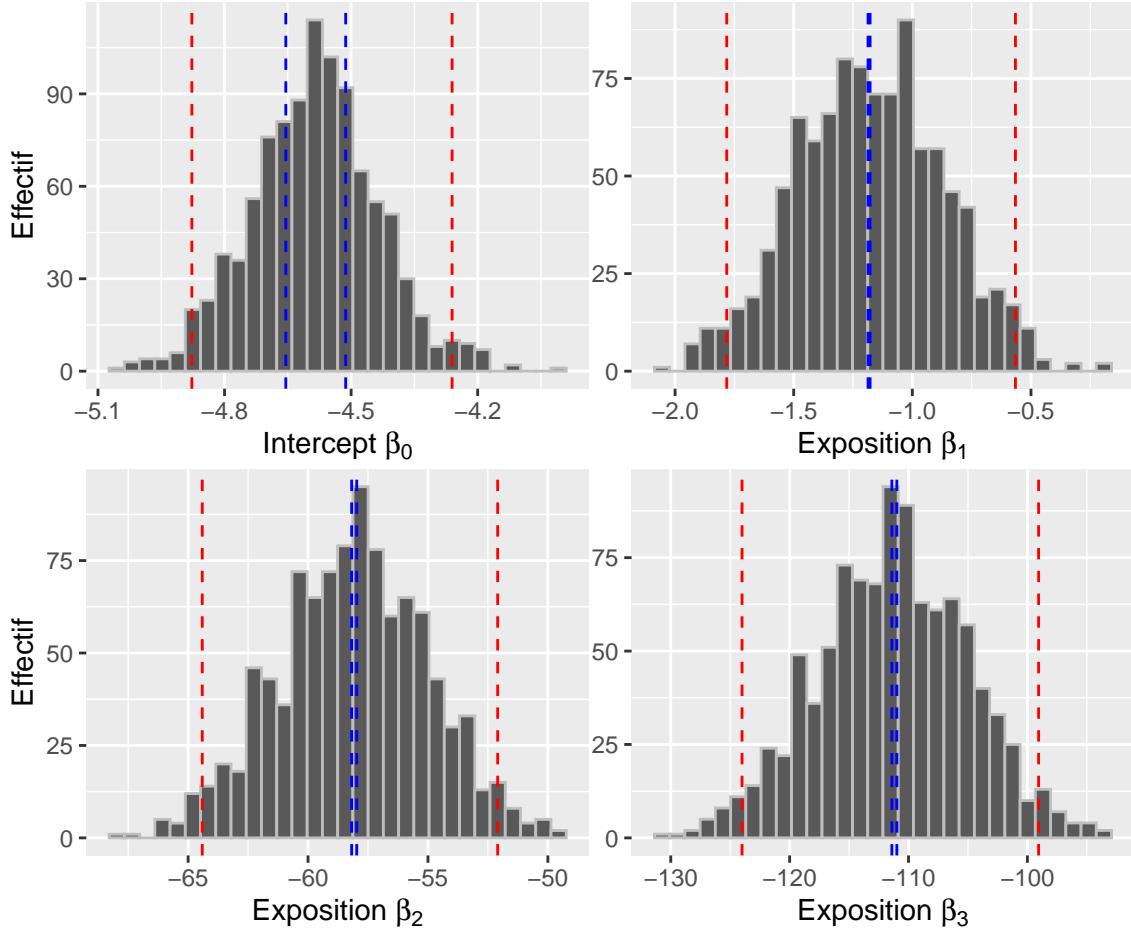


Figure 14: Intervalles de confiance dans le cas d'une loi normale multivariée sans double bi-exponentielle.

2.4.2 Pourcentage du taux normal d'émission de buzz

Il est rassurant de voir que les intervalles de confiance des coefficients d'exposition obtenus par Monte Carlo restent raisonnablement petits, cependant comme évoqué précédemment, il n'est pas possible de lier une augmentation ou une diminution du taux d'émission de buzz en fonction de la distance avec le bateau à partir de leur valeur. C'est pourquoi nous avions représenté graphiquement cette évolution sur la figure 4, mais cette visualisation ne proposait pas d'intervalle de confiance ou de bande de prédition.

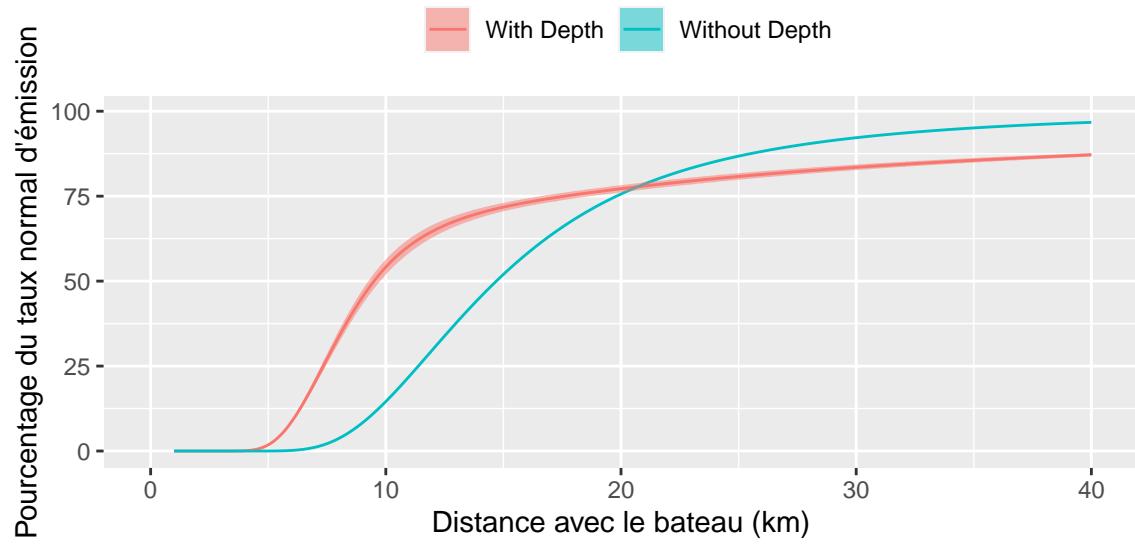


Figure 15: Pourcentage du taux normal d'émission de buzz selon la distance au bateau.