



**Master Mathématiques,
Spécialité Mathématiques et applications,
Parcours Statistiques et Recherche Opérationnelle**

Rapport de stage pour l'obtention
de la 1^{ère} année de Master

Estimation des sources d'erreurs sur la mesure des teneurs en contaminants dans les sols du Réseau de Mesure de Qualité des Sols par analyse de variance hiérarchique

Hakima Boukir

Mai-août, 2012

Maîtres de stage : Nicolas Saby, Sophie Jacquot,
Line Boulonne, Claudy Jolivet
INRA, Orléans, unité InfoSol



Institut National de la Recherche Agronomique

Mes remerciements

Je tiens en premier lieu à remercier Marion Bardy, directrice de l'unité InfoSol d'Orléans, pour m'avoir offert l'opportunité d'effectuer mon stage au sein de l'INRA. J'adresse par ailleurs mes remerciements à Nicolas Saby pour avoir pris de son temps afin de m'initier aux dédales de la statistique et ce toujours avec patience et pédagogie.

Je remercie également Sophie Jacquot, Line Boulonne et Claudy Jolivet pour leur grande disponibilité et leur soutien tout au long de ce stage.

Enfin, je remercie l'ensemble du personnel de l'INRA et en particulier Jennifer Granja, Robin Dagois, Florent Millet, Chloé Swiderski, Kevin Vaysse et Fanny Collard pour les moments mémorables qui m'ont été donnés de vivre durant ces quelques mois et sans quoi ma formation n'aurait été que moins passionnante.

Table des matières

Introduction	3
1. Matériel et méthodes	5
1.1. Schéma d'échantillonnage et préparation des échantillons	5
1.2. Schéma de mesure de répétabilité des teneurs en éléments	6
1.3. Schéma de répétabilité des densités apparentes et des teneurs en éléments grossiers	6
1.4. Calcul du stock de chaque élément	7
1.5. Modèles linéaires mixte hiérarchiques de calcul de la variance des teneurs en éléments	7
1.5.1. <i>Expression du modèle linéaire mixte hiérarchique</i>	7
1.5.2. <i>Ajustement du modèle par recherche du maximum de vraisemblance</i>	8
1.6. Outils graphiques	8
1.6.1. <i>Les boxplots</i>	8
1.6.2. <i>Le graphique Q-Q (Quantile-Quantile)</i>	8
1.7. Simulation d'un jeu de données associé aux modèle mixte hiérarchique pour différents schéma	8
1.8. Transformation de Box-Cox	10
2. Résultats et discussion	10
2.1. Les simulations	10
2.2. Ajustement du modèle linéaire mixte hiérarchique pour les données du RMQS	14
2.2.1. <i>Approche descriptive</i>	14
2.2.2. <i>Comparaison des différents modèles de calcul de la variance des teneurs en éléments</i> 17	17
2.2.3. <i>Transformation des variables</i>	19
2.3. Composants de la variance	21
Conclusion	22
Bibliographie	24
Annexe 1 : Statistiques descriptives des sites de répétabilité pour la couche de surface.	25
Annexe 2 : Statistiques descriptives des sites de répétabilité pour la couche de sub- surface	27
Annexe3 : Histogrammes de tous les éléments pour la couche de surface	28
Annexe 4 : Boxplots des teneurs en éléments par site de répétabilité pour la couche de surface	30
Annexe 5 : Boxplots des teneurs en éléments par site de répétabilité pour la couche de sub-surface	35
Annexe 6 : Tableau de comparaisons des modèles mixtes hiérarchiques	41
Annexe 7 : Boxplots centrés des teneurs en élément par site de répétabilité pour la couche de surface	43
Annexe 8 : Boxplots centrés des teneurs en élément par site de répétabilité pour la couche de sub-surface	48
Annexe 9 : Composants de la variance estimés du modèle qui minimise l'AIC pour la couche de surface	53
Annexe 10 : Composants de la variance estimés du modèle qui minimise l'AIC pour la couche de sub-surface	53

Résumé

Les erreurs associées aux estimations des teneurs en différents éléments dans les sols sont rarement quantifiées bien qu'elles puissent empêcher l'obtention de résultats significatifs. Cette étude vise à analyser les sources d'erreurs liées aux mesures des teneurs en différents contaminants dans les sols étudiés à travers le Réseau de Mesures de Qualité des Sols. Dans ce cadre, un jeu de données suivant un schéma de répétabilité a été mis en place. Les sources d'erreur sont estimées par ajustement d'un modèle linéaire mixte avec effets aléatoires emboîtés. Ces effets, correspondant aux composants de la variance de chaque étape analytique, sont estimés par recherche du maximum de vraisemblance (REML). Le modèle a été ajusté sur des données simulées et réelles. Les résultats des simulations montrent l'absence de biais dans l'estimation des paramètres pour la structure du schéma RMQS. Cependant, la précision reste faible ne permettant pas de distinguer les différentes sources d'erreurs. L'application de ce modèle aux données du RMQS montre la difficulté à trouver un modèle respectant les hypothèses statistiques.

Abstract

Uncertainties in soil properties assessments are rarely quantified even though they are critical in determining the significance of the results. This study aims to examine sources of measurements uncertainties in soil contaminants through the "Réseau de Mesures de Qualité des Sols". As part of the RMQS, a dataset of measurements has been obtained according to a repeatable pattern. Sources of uncertainties were estimated by fitting mixed model with random nested effects. These effects, which are the components of variance for each analytical step, were estimated by residual maximum likelihood (REML) method. The model was fitted on simulated and real dataset. Simulation results showed that parameters estimation for the RMQS scheme is unbiased. However, because of weakness in precision we can't determine different sources of uncertainties. The application of model to the RMQS dataset shows how difficult it is to find a model who meets statistic assumptions.

Introduction

Les sols forment une ressource naturelle non renouvelable. La surveillance de l'évolution de leur qualité constitue donc un enjeu environnemental important. Le Réseau de Mesure de Qualité des sols (RMQS) est alors mis en place en France par l'intermédiaire du Groupement d'Intérêt Scientifique du Sol (GIS Sol). Il s'agit d'un outil mis en œuvre afin de permettre d'évaluer et de suivre la qualité des sols. Le RMQS est une grille couvrant la France (les DOM compris) de 2200 mailles de 16 km² de côté. Ainsi, quatre campagnes de prélèvement sont prévues, avec un pas de temps d'environ 10 ans. La première campagne de prélèvement est entièrement terminée.

Il est donc nécessaire de connaître l'évolution des teneurs en contaminants dans les sols au cours du temps. La détection et le calcul des incertitudes permettent alors de s'assurer des évolutions constatées et de l'obtention de résultats significatifs. Il existe plusieurs types d'incertitudes : des incertitudes peuvent être liées au prélèvement et à la variabilité spatiale, d'autres à la réduction de l'échantillon et aux mesures en laboratoire (Goidts *et al*, 2009).

La mesure de ces incertitudes repose sur la quantification et la séparation des erreurs associées à chaque étape analytique (Rawlins *et al*, 2009). Parallèlement, un plan d'échantillonnage emboîté et adapté aux différentes étapes analytique de mesures des teneurs en contaminant est ainsi mis en place dans le cadre du programme RMQS. Les sources d'erreurs sont alors mesurées au moyen de modèles mixtes hiérarchiques appliqués sur le jeu de données issu du schéma de répétabilité établie par le RMQS (Webster *et al*, 2006 ; Lark *et al*, 2011). Pour s'assurer de l'efficacité de l'estimation des schémas de répétabilité mis en place, les modèles sont également ajustés sur des données simulées (Papritz *et al*, 2011).

Ainsi, dans ce rapport nous présenterons la structure du schéma de répétabilité mis en place au sein du programme RMQS. Nous simulerons ensuite des jeux de données associées au modèle mixte hiérarchique pour vérifier l'absence de biais dans les estimations. Enfin, nous ajusterons le modèle avec effets aléatoires imbriqués et estimerons les variances de chaque étape analytique.

1. Matériel et méthodes

1.1. Schéma d'échantillonnage et préparation des échantillons

Une surface d'échantillonnage suivant un plan grille non alignée de 20 m par 20 m, découpée en 100 carrés élémentaires est installée au centre de chaque maille du réseau RMQS. Lors de chacune des quatre campagnes prévues, un quart de ces surfaces élémentaires est échantillonné. Ainsi, au cours de la première campagne, les carrés 1 sont prélevés puis les carrés 2,3 et 4 lors des campagnes suivantes (figure1). Les 25 prélèvements ainsi échantillonnés à la tarière¹ sont mélangés pour constituer un échantillon composite. Pour décrire le sol et déterminer les profondeurs d'échantillonnage des prélèvements composites, une fosse pédologique est creusée à 5 m de la surface d'échantillonnage. Les prélèvements sont effectués dans la couche supérieure du sol et dans la couche sous-jacente, lorsqu'elle est présente. Le premier prélèvement est réalisé entre 0 et 30 cm de profondeur si le sol n'est pas cultivé ou sur l'épaisseur de la couche travaillée. Le deuxième prélèvement est réalisé entre 30 et 50 cm ou sous la couche travaillée jusqu'à 50 cm. Ces deux couches de sol étant de nature différente que ce soit en termes de densité apparente ou de teneur en éléments chimiques, elles sont traitées séparément dans l'analyse présentée ici.

Plusieurs étapes réalisées au Conservatoire de l'INRA d'Orléans sont nécessaires à la préparation des échantillons. Les échantillons alors constitués sont transmis au Laboratoire des sols de l'INRA d'Arras pour le dosage des éléments chimiques. Les éléments chimiques considérés ici sont l'azote, le cadmium, le chrome, le cobalt, le cuivre, le molybdène, le nickel, le plomb, le thallium et enfin le zinc. Les échantillons sont tout d'abord séchés à l'air libre à l'abri de toutes contaminations. Puis, l'échantillon est quarté afin de préparer 500 grammes d'échantillons représentatifs, émottés et tamisés à 2 mm. Les échantillons sont alors envoyés au laboratoire d'analyses pour le dosage. La première étape du dosage consiste au prélèvement d'une aliquote de 30 grammes. Lors de la seconde étape, l'aliquote est broyée et tamisée à 250 µm avec une prise d'essai de 50 mg. Enfin, cette prise d'essai est dosée par combustion sèche. Au total, quatre étapes sont nécessaires aux mesures des teneurs en éléments chimiques d'un sol : 1-quartage et préparation de 500g de 2mm, 2-quartage de 30 g et broyage à 250 µm, 3- prise d'essai de 50 mg, et 4- dosage par combustion sèche.

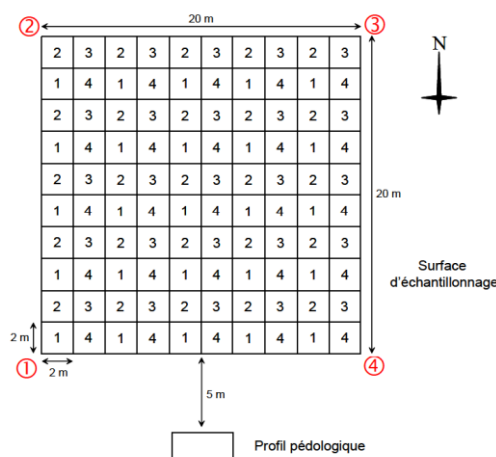


Figure 1: Dispositif de prélèvement et de description du RMQS

¹ Une tarière est un outil permettant de faire des sondages pédologiques

1.2. Schéma de mesure de répétabilité des teneurs en éléments

Afin d'analyser et de discriminer quantitativement les différentes sources de variabilité dans la mesure des teneurs en éléments, un jeu de données A correspondant au schéma de répétabilité de la figure 2 a été collecté sur neuf sites du réseau RMQS répartis en France. Les sites sélectionnés présentent des situations contrastées de types de sols et d'occupation. Le protocole consiste au prélèvement de tous les carrés afin de former quatre composites par couche (supérieure et sous-jacente) de chaque site. Un des quatre composites est ensuite subdivisé en cinq sous-échantillons pour la première étape. Un de ces sous-échantillons est lui-même subdivisé pour la deuxième étape, et ainsi de suite jusqu'à la dernière étape. Parmi les sites sélectionnés quatre ne présentent pas l'étape de broyage à 250 µm et de prise d'essai.

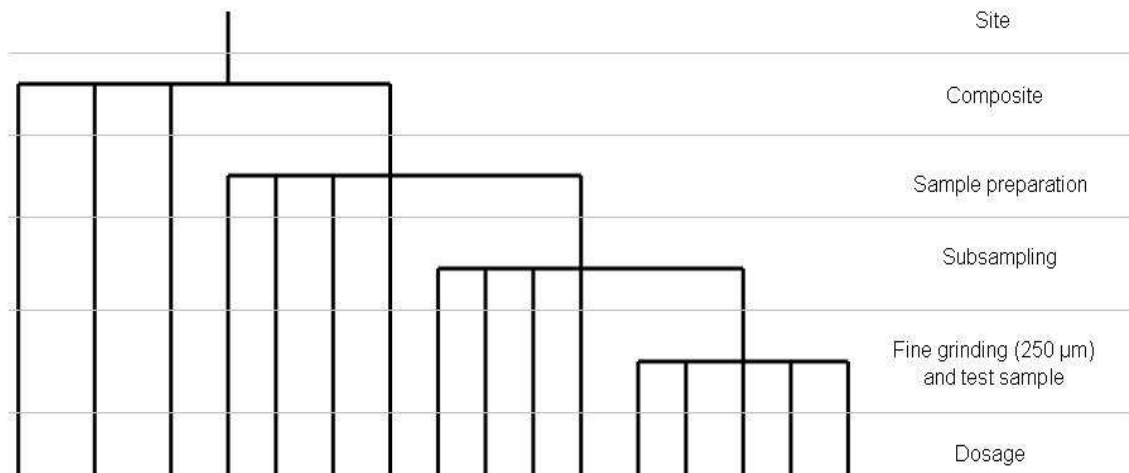


Figure 2: Schéma de répétabilité pour les mesures de teneurs en éléments

1.3. Schéma de répétabilité des densités apparentes et des teneurs en éléments grossiers

Sur chaque site RMQS, des mesures de densité apparente et des teneurs en éléments grossiers sont effectuées en parallèle des prélèvements pour la mesure des teneurs en éléments chimiques à partir de prélèvements volumétriques.

Un second jeu de données B, correspondant au schéma de répétabilité pour les mesures de densité apparente et de la teneur en éléments grossiers a ainsi été mis en place. Ces données ont été utilisées pour évaluer dans un premier temps les incertitudes associées à la détermination de la proportion relative en éléments grossiers et à la mesure de la densité apparente pour l'évaluation des erreurs sur les stocks en élément chimique. Ces deux variables ont été mesurées sur 225 sites du RMQS avec deux à trois répétitions à la même profondeur, dans les couches de sol. Pour les mesures de densité apparente, deux méthodes différentes ont été utilisées. La première méthode appelée méthode du cylindre, consiste à peser un échantillon de sol prélevé à l'aide d'un cylindre de 500 ml. La seconde dite méthode à l'eau consiste à creuser une cavité sur une surface aplanie du sol, à en recueillir la totalité de la terre afin de la peser, puis à placer un sac plastique étanche afin de le remplir d'eau dont on mesurera le volume.

1.4. Calcul du stock de chaque élément

Le stock de chacun des éléments chimiques dans le sol représente la masse de l'élément chimique rapportée à une unité de surface, pour une épaisseur de sol donnée. Il est calculé par le biais de l'équation suivante :

$$stock = \frac{d \times E \times BD \times [1 - RM]}{100} \quad (1)$$

Où *stock* désigne le stock de l'élément en tonne d'élément par hectare (t E.ha⁻¹), *d* l'épaisseur de couche considérée (m), *RM* la masse volumique apparente du sol (kg.m⁻³), *BD* la proportion massique relative d'éléments grossiers (sans dimension) et la teneur en élément dans le sol (g E.kg⁻¹). Ici, *E* doit être remplacé par l'élément chimique calculée, autrement dit par *N*(azote), *Cd*(cadmium), *Cr*(chrome), *Co*(cobalt), *Cu*(cuivre), *Mo*(molybdène), *Ni*(nickel), *Pb*(plomb), *Tl*(thallium) ou *Zn*(Zinc).

1.5. Modèles linéaires mixte hiérarchiques de calcul de la variance des teneurs en éléments

1.5.1. Expression du modèle linéaire mixte hiérarchique

Pour estimer les différentes sources de variabilité des mesures des teneurs en élément aux différentes étapes analytiques, il est nécessaire de prendre en compte le fait que chaque variance d'une étape analytique donnée intègre la variance des étapes précédentes. Ainsi, des modèles mixtes hiérarchiques sont mis en œuvre et permettent d'extraire la variance propre d'une étape analytique. Un modèle mixte permet d'intégrer à la fois des effets fixes mais aussi des effets aléatoires. Les effets fixes permettent ici d'expliquer les moyennes de la variable par effet. Les effets aléatoires emboîtés dans le cas d'un modèle hiérarchique permettent d'expliquer la variance autour de cette moyenne (les différentes étapes de mesure). Considérons un sous-échantillon *k* issu du sous-échantillonnage *g* de l'étape d'émottage et de tamisage *h* du composite *j* dans le site *i*. Le modèle associé au calcul de la variance de la teneur en élément dans le sol peut alors s'écrire pour un sous-échantillon :

$$Z_{ijhpk} = \mathbf{x}_{ijhpk}^T \boldsymbol{\beta} + A_i + B_{ij} + C_{ijh} + D_{ijhg} + \varepsilon_{ijhpk} \quad (2)$$

Avec \mathbf{T} qui signifie la transposée ; \mathbf{x}_{ijhpk} et $\boldsymbol{\beta}$ sont les vecteurs des covariables et des coefficients des régressions représentant les effets fixes, lorsqu'il n'y a pas d'effets fixes \mathbf{x} est un vecteur uniquement constitué de 1, ε_{ijhpk} est l'erreur résiduelle représentant la déviation de la valeur observée au dernier niveau d'analyse ; et les termes restants désignent les effets aléatoires emboîtés : A_i variable aléatoire représentant le composite *i*, B_{ij} la différence entre la moyenne du composite *j* dans le site *i*, C_{ijh} la différence entre la moyenne de l'étape *h* dans le composite *j* dans le site *i* et la moyenne du composite *j* et ainsi de suite. Le modèle repose sur l'hypothèse que le vecteur des observations \mathbf{z} est une variable aléatoire de distribution :

$$\mathbf{z} \sim \mathcal{N}(\mathbf{X}^T \boldsymbol{\beta}, \mathbf{V}),$$

où \mathbf{X} est la matrice $n \times p$ des effets fixes dont les vecteurs \mathbf{x}_{ijhpk} de l'équation (2) sont les lignes. \mathbf{V} est la matrice de variance covariance définie par :

$$\mathbf{V} = \sum_{l=1}^5 \sigma_l^2 \mathbf{U}_l \mathbf{U}_l^T$$

où les $\sigma_1, \dots, \sigma_5$ sont les variances des effets aléatoires et \mathbf{U}_l est une matrice qui renseigne sur la structure du schéma de prélèvement. Si on a n observations, et au niveau l il y a n_k unité, alors \mathbf{U}_l est une matrice $n \times n_k$. Si l'observation g est dans l'unité j au niveau 1, alors l'élément dans la colonne j de la ligne g prend la valeur 1 et le reste 0. Le log de la vraisemblance peut alors s'écrire :

$$l_R = -\frac{1}{2}(\ln V + \ln |\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}| + \mathbf{z}^T \mathbf{P} \mathbf{z})$$

où \mathbf{P} est égale à :

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1}.$$

1.5.2. Ajustement du modèle par recherche du maximum de vraisemblance

Les composants de la variance pour chacun des effets aléatoires ainsi que les effets fixes sont estimés par recherche du maximum de vraisemblance restreint (*residual maximum likelihood*).

Nous avons testé dans cette étude plusieurs covariables pour tenter de rendre les effets aléatoires centrés. La comparaison des différents effets fixes testés se fait par comparaison des AIC. Pour cela, il est nécessaire de d'ajuster les modèles par maximisation de la vraisemblance.

1.6. Outils graphiques

1.6.1. Les boxplots

La boîte à moustache (boxplot en anglais) est un outil graphique représentant une distribution empirique à l'aide de plusieurs paramètres : la médiane, le 1^{er} et le 3^{ème} quartile. Le boxplot peut décrire une variable seule ou par groupe. La boîte est ainsi délimitée en haut par le 3^{ème} quartile et en bas par le 1^{er} quartile, à l'intérieur de la boîte une droite représente la médiane (2^{ème} quartile). Les valeurs minimales et maximales sont également dessinées : les segments ajoutés mènent aux valeurs extrêmes qui sont soit le 1^{er} et le 9^{ème} décile ou alors le minimum et le maximum.

1.6.2. Le graphique Q-Q (Quantile-Quantile)

Construire un graphique Q-Q (Quantile-Quantile) également appelé diagramme quantile-quantile (Q-Q plot en anglais) est une manière d'évaluer visuellement la normalité d'une distribution. Ce graphique consiste à représenter conjointement la distribution empirique des résidus avec la distribution théorique qu'ils sont censées suivre, en l'occurrence, la distribution normale. En abscisse se trouvent les quantiles théoriques et en ordonnée les quantiles observées. Si les deux distributions sont linéairement liées leur représentation conjointe doit s'aligner sur une droite. Dans notre étude, ce diagramme quantile-quantile permet de vérifier la normalité des résidus du modèle linéaire mixte.

1.7. Simulation d'un jeu de données associé aux modèle mixte hiérarchique pour différents schémas

Un schéma de répétabilité est dit équilibré si chaque étage du schéma est constitué du même nombre de points. Pour ce type de schéma contrairement aux schémas déséquilibrés, les estimateurs par les moindres carrés conviennent. Pourtant, ce type de schéma nécessite très vite un grand nombre de valeurs. Lark (2012) montre que l'on peut contourner ce problème en ayant recours à l'estimation par maximisation de la vraisemblance sur des schémas déséquilibrés. L'estimateur par maximisation de la vraisemblance est en effet asymptotiquement non biaisé et efficace pour des schémas déséquilibrés. Pour cela, il propose une méthode d'optimisation permettant de définir le meilleur schéma. Cependant, dans le cadre de notre étude, le schéma est fixé et très déséquilibré. Comme proposé par Papritz et al (2011), nous avons voulu tester l'efficacité de différents schémas à l'aide de techniques de *Monte-Carlo*. Nous avons ainsi simulé plusieurs jeux de données issues du modèle mixte hiérarchique de l'équation (2) pour un schéma équilibré à 2 et 4 individus par étage et pour le schéma du RMQS. L'objectif est alors de comparer les estimations fournies par le modèle et les variances imposées. Les variances σ^2_{site} , σ^2_{comp} , σ^2_{id2} , σ^2_{id3} , σ^2_{res} sont les composants de la variance. Par exemple, pour simuler une réalisation de 16 valeurs issues d'un site selon un schéma équilibré avec 2 sites par étage, on tire un nombre aléatoire a_i issu d'une distribution normale $\mathcal{N}(0, \sigma^2_{\text{site}})$. Ensuite, on tire deux nombres aléatoires b_{i1}, b_{i2} , indépendants de loi $\mathcal{N}(0, \sigma^2_{\text{comp}})$, quatre nombres $c_{i11}, c_{i12}, c_{i21}, c_{i22}$ de loi $\mathcal{N}(0, \sigma^2_{\text{id2}})$, huit nombres $d_{i111}, d_{i112}, d_{i121}, \dots, d_{i222}$ de loi $\mathcal{N}(0, \sigma^2_{\text{id3}})$ et enfin 16 erreurs résiduelles $e_{i1111}, \dots, e_{i2222}$ de loi $\mathcal{N}(0, \sigma^2_{\text{res}})$. Les 16 données simulées $z_{i1111}, \dots, z_{i2222}$ sont alors bien obtenues à partir de l'équation

$$Z_{ijhgk} = x_{ijhgk}\beta + A_i + B_{ij} + C_{ijh} + D_{ijhg} + \varepsilon_{ijhgk}$$

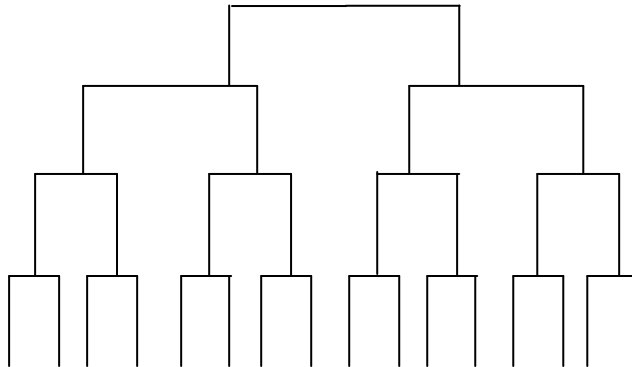


Figure 3 : Schéma de répétabilité pour un modèle équilibré à 2 points par étage

L'ajustement des modèles sur la simulation peut être évalué par l'erreur quadratique moyenne (*root mean square error*)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\sigma_i^2 - \sigma^2)^2} \quad (3)$$

où σ^2 désigne un composant de la variance, σ_i^2 son estimation calculée à partir de la i ème réplication et n le nombre de réplifications.

1.8. Transformation de Box-Cox

L'ajustement d'un modèle linéaire mixte hiérarchique comme la plupart des modèles statistiques repose en partie sur la normalité de la distribution étudiée. Ici, les variables étudiées sont les teneurs en éléments dans le sol. Ainsi, on cherche à transformer les variables de départ afin de se rapprocher d'une distribution gaussienne. L'une des transformations possible est la transformation non-linéaire de Box-Cox. Elle est définie comme :

$$Z(x, \lambda) = \begin{cases} \frac{x^\lambda - 1}{\lambda}, & \text{si } \lambda \neq 0 \\ \log(x), & \text{si } \lambda = 0 \end{cases}$$

où Z désigne la variable étudiée et x est positif.

2. Résultats et discussion

2.1. Les simulations

Nous avons simulé trois jeux de données issus du modèle linéaire mixte hiérarchique composés chacun de 1000 réplifications de la réalisation. Le premier schéma simulé correspond à un jeu de données équilibré avec 2 points par étage, le deuxième schéma équilibré est constitué de 4 points par étage. Enfin, le schéma déséquilibré de mesures de répétabilité des teneurs en éléments du RMQS est simulé.

On simule pour chacun de ces schémas 1000 jeux de données en imposant un σ_i (i = site, comp, id2, id3, res) composant de la variance. L'objectif est alors de comparer les variances estimées pour chacun des modèles avec les variances imposées et de justifier l'efficacité du schéma déséquilibré établie pour le RMQS.

Nous avons ainsi simulé 1000 réplifications de la réalisation du modèle pour 9 sites en imposant comme variances $\mathbf{v}_1 = (12, 0.3, 0.3, 0.4, 0.3)$, variances estimées pour le nickel puis on impose arbitrairement $\mathbf{v}_2 = (5, 8, 2, 6, 1)$. Nous avons choisi 2 ensembles de variances contrastés afin de tester les modèles dans des conditions différentes. Ainsi, sur les figures 4 et 5, l'axe des abscisses correspond aux composants de la variance σ_{comp} (B), σ_{id2} (C), σ_{id3} (D) et σ_{res} (e). Sur l'axe des ordonnées on retrouve les valeurs des variances. On représente la cible, à savoir la variance imposée par un cercle et la moyenne de la variance estimée par simulation pour chaque composante par une croix. Enfin, le 1^{er} et le 9^{ème} décile sont dessinés au moyen d'un trait. Avec \mathbf{v}_1 et pour les deux schémas équilibrés, les composants de la variance sont tous bien estimés et leurs estimations semblent peu biaisées. Cependant, l'incertitude est plus grande pour l'estimation du composant de la variance associé à l'étape du composite avec des intervalles de confiance plus larges. Tous les schémas sont efficaces mais le schéma équilibré à 4 points semble être le moins biaisé. Cependant, contrairement au schéma déséquilibré à 4 points, le schéma RMQS nécessite dans sa construction moins de points et est donc moins coûteux en calcul. La simulation permet ainsi de justifier l'efficacité du schéma de mesure de répétabilité du RMQS. Avec \mathbf{v}_2 , tous les schémas simulés estiment correctement chacun des composants de la variance y compris le composant correspondant à l'effet composite. On observe cependant des intervalles d'estimation autour de la moyenne plus large pour σ_{comp} et σ_{id2} .

On compare ensuite les erreurs quadratiques moyennes pour les composants de la variance σ_{comp} (comp), σ_{id2} (id2), σ_{id3} (id3) et σ_{res} (sigma) de chacun des schémas. Sur les figures 6 et 7 sont tracées les erreurs quadratiques moyennes pour chaque simulation et par composante de la variance.

Pour les deux variances imposées et pour chaque schéma, l'erreur quadratique moyenne est la plus forte pour l'estimation du composant de la variance correspondant à l'étape du composite et diminue de façon décroissante jusqu'à la variance des résidus. Avec le schéma déséquilibré à 4 points on a les erreurs les plus faibles par rapport au schéma RMQS. En effet, pour le schéma RMQS on a par exemple, une erreur de 50% contre seulement 20% pour l'estimation de σ_{comp} . Pour l'estimation de la variance résiduelle, on n'a plus qu'une erreur de 2% pour le schéma équilibré à 4 points contre 10% pour le schéma RMQS. Avec v_2 , on retrouve les plus faibles erreurs quadratiques moyennes pour chaque composant de la variance avec le schéma équilibré à 4 points par étage.

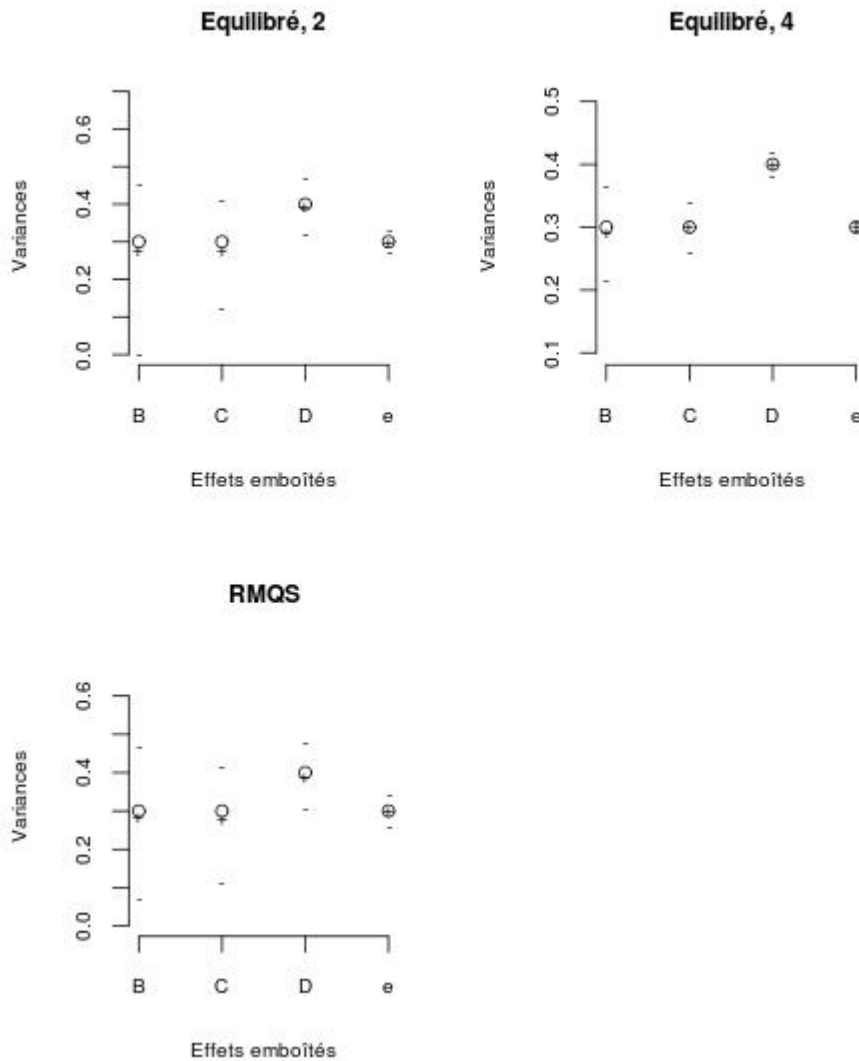


Figure 4 : Comparaison des moyennes des variances estimées pour 1000 répliques avec la variance imposée $v_1 = (12, 0.3, 0.3, 0.4, 0.3)$. Variance imposée et moyenne de la variance estimée par simulation pour chaque composante représentée respectivement par un cercle et une croix. Le 1^{er} et 9^{ème} décile sont représentés par un trait.

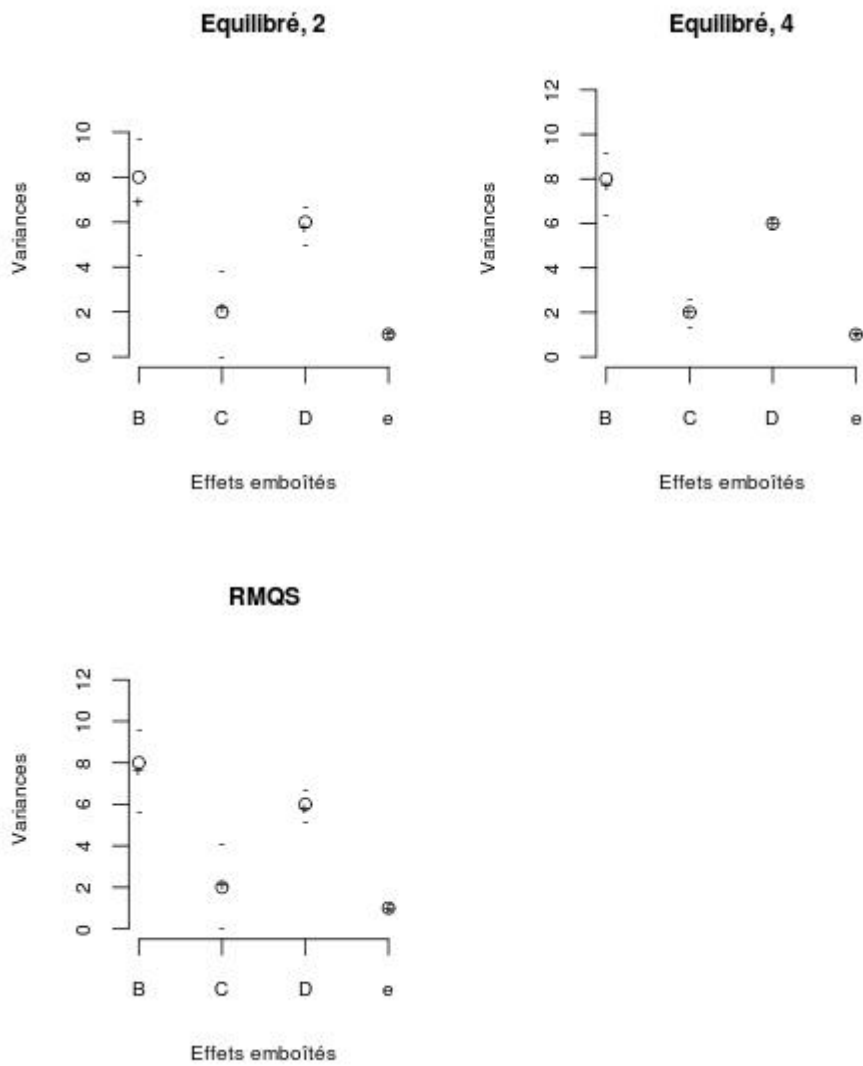


Figure 5 : Comparaison des moyennes des variances estimées pour 1000 réplicas avec la variance imposée $v_2 = (5, 8, 2, 6, 1)$. Variance imposée et moyenne de la variance estimée par simulation pour chaque composante représentée respectivement par un cercle et une croix. Le 1^{er} et 9^{ème} décile représentés par un trait.

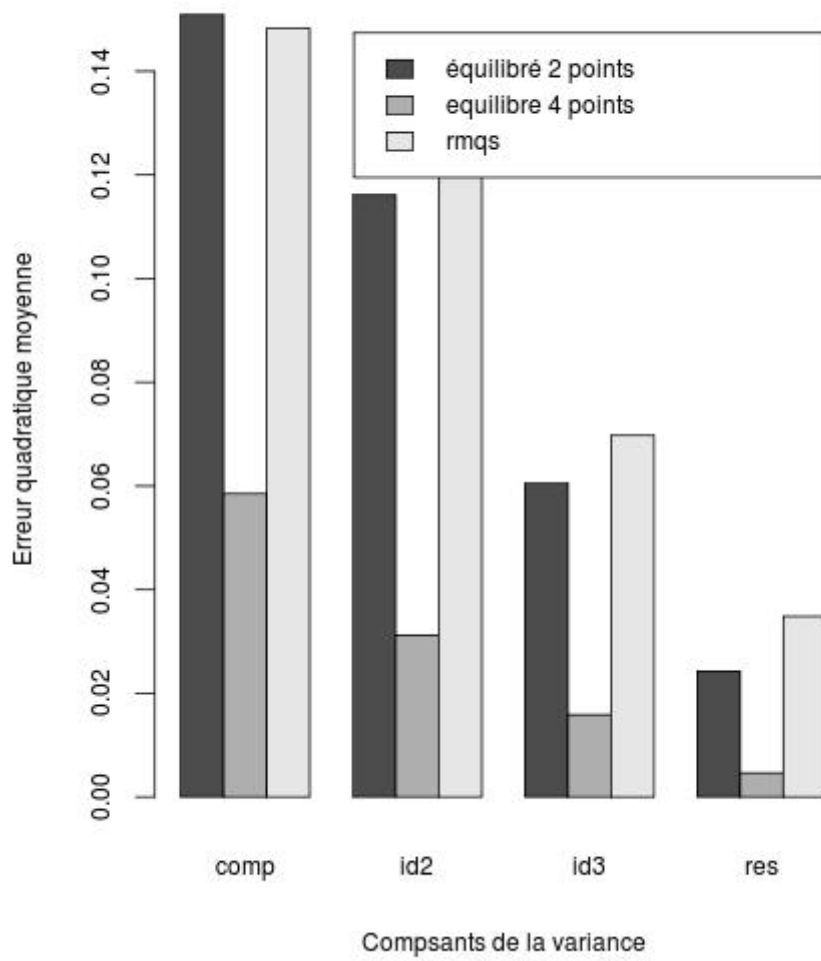


Figure 6 : Erreur quadratique moyenne par composant de la variance pour les trois schémas avec $v_1 = (12, 0.3, 0.3, 0.4, 0.3)$.

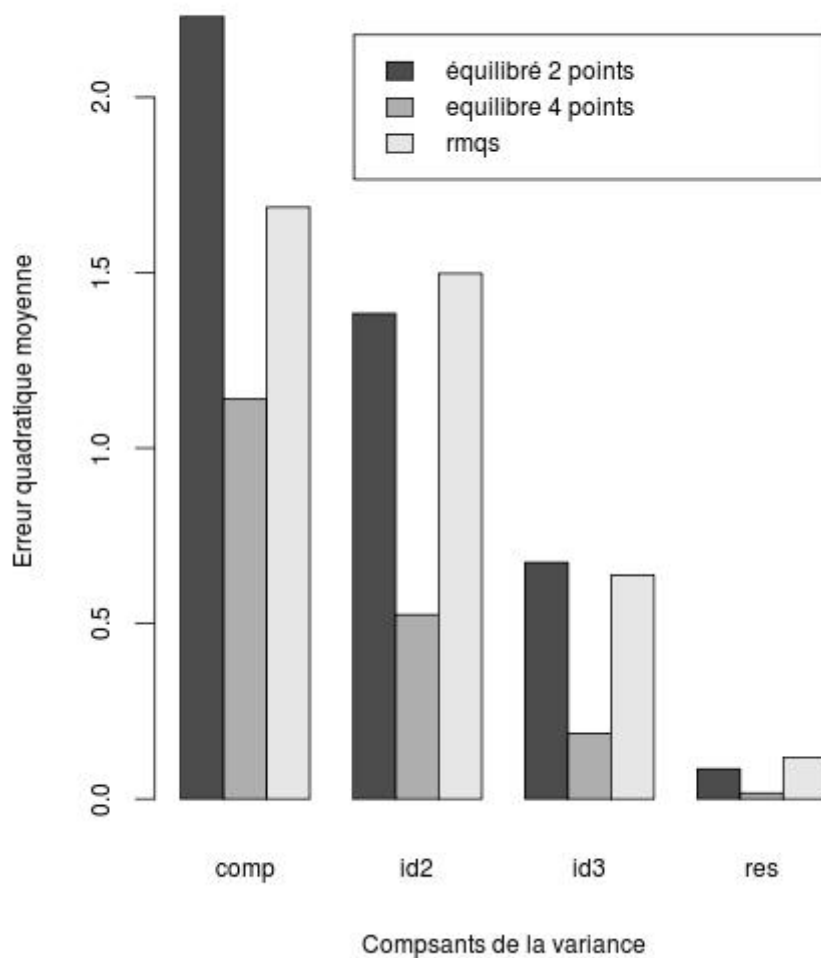


Figure 7 : Erreur quadratique moyenne par composant de la variance pour les trois schémas avec $v_2 = (5, 8, 2, 6, 1)$

2.2. Ajustement du modèle linéaire mixte hiérarchique pour les données du RMQS

Dans la suite, nous focaliserons l'étude sur 4 exemples, Ni et Mo pour la couche de surface, Pb et Zn pour la couche de sub-surface, pour simplifier la compréhension des résultats.

2.2.1. Approche descriptive

2.2.1.1. Approche descriptive pour la couche de surface

On considère le nickel et le molybdène, deux des dix variables étudiées et on souhaite améliorer le modèle mixte hiérarchique de départ correspondant à l'équation (2) afin d'évaluer les variabilités de leur teneur dans le sol. Une première approche consiste à observer les statistiques descriptives des teneurs en nickel et en molybdène par site pour la couche de surface (annexe 1). Ainsi, les moyennes par site des teneurs de chaque élément trace et leur variance sont présentées. Les valeurs minima et maxima permettent de décrire l'étendue des valeurs. Cette première approche descriptive permet d'appréhender les différentes sources de

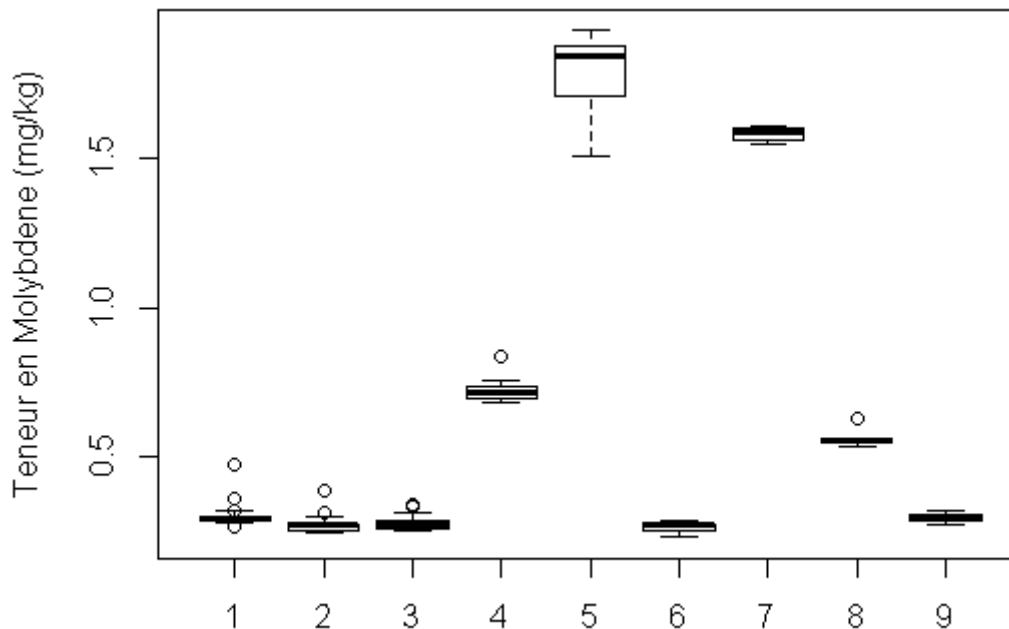
variabilité et d'ajuster un modèle mixte hiérarchique qui intégrerait les variables les plus pertinentes.

Au sein d'un même site, les variances sont assez élevées pour le nickel et beaucoup plus faible pour le molybdène. Le comportement des valeurs moyennes de chaque élément par site permet d'avoir une première analyse de l'influence du site sur la teneur en élément chimique. Ainsi, la teneur en élément varie d'un site à l'autre. Les différences entre sites sont plus grandes pour le nickel et moins importantes pour le molybdène. D'autre part, les moyennes les plus élevées se retrouvent à la fois pour le nickel et le molybdène au sein du même site (site 11151).

Tableau 1 : Tableau de correspondances des numéros de site

Site	689	801	899	1208	11151	10797	1326	269	989
Numéro Site	1	2	3	4	5	6	7	8	9

La figure 8 représente les teneurs en molybdène et en nickel par site pour la couche de surface. Le tableau 1 permet alors de retrouver la correspondance des sites. De ce fait, l'analyse des teneurs par site confirme un effet site mais d'importance variable. Ainsi, le site semble être une variable beaucoup plus discriminante pour le nickel que pour le molybdène.



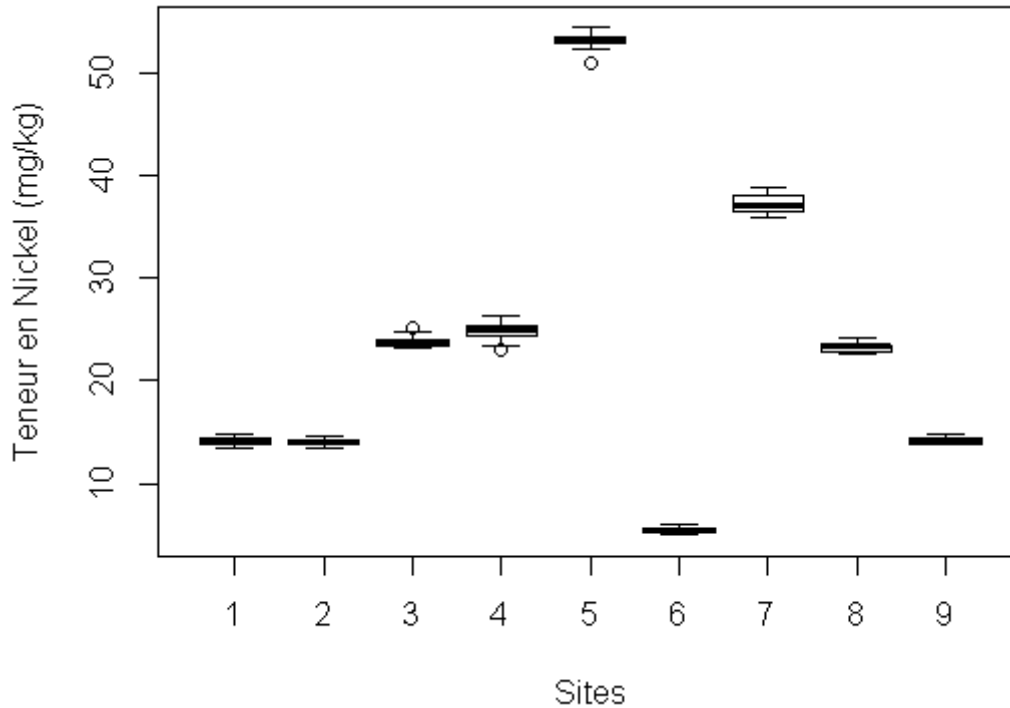


Figure 8 : Boxplots des teneurs en molybdène et en nickel par site pour la couche de surface

2.2.1.2. *Approche descriptive pour la couche de sub-surface*

On observe les statistiques descriptives des teneurs dans le sol par site pour le plomb et le zinc dans la couche de sub-surface (annexe 2). Les variances au sein d'un même site sont élevées pour le zinc et plus faible pour le plomb. Les valeurs moyennes de la teneur en zinc varient énormément en fonction du site. Un effet « site » semble ainsi judicieux pour le modèle de calcul de la variance des teneurs en zinc dans le sol pour la couche de sub-surface. Au contraire, pour le plomb le comportement des moyennes par site des teneurs ne s'avère pas significatif pour justifier un effet « site » dans le modèle associé au calcul des variances des teneurs en plomb.

Les représentations des teneurs en zinc et en plomb par site (figure 9) semblent confirmer les hypothèses formulées juste avant. En effet, un effet « site » semble plus discriminant pour le zinc que pour le plomb. De plus, on observe une grande disparité selon les sites dans les teneurs en zinc, ce qui n'est pas le cas pour le plomb.

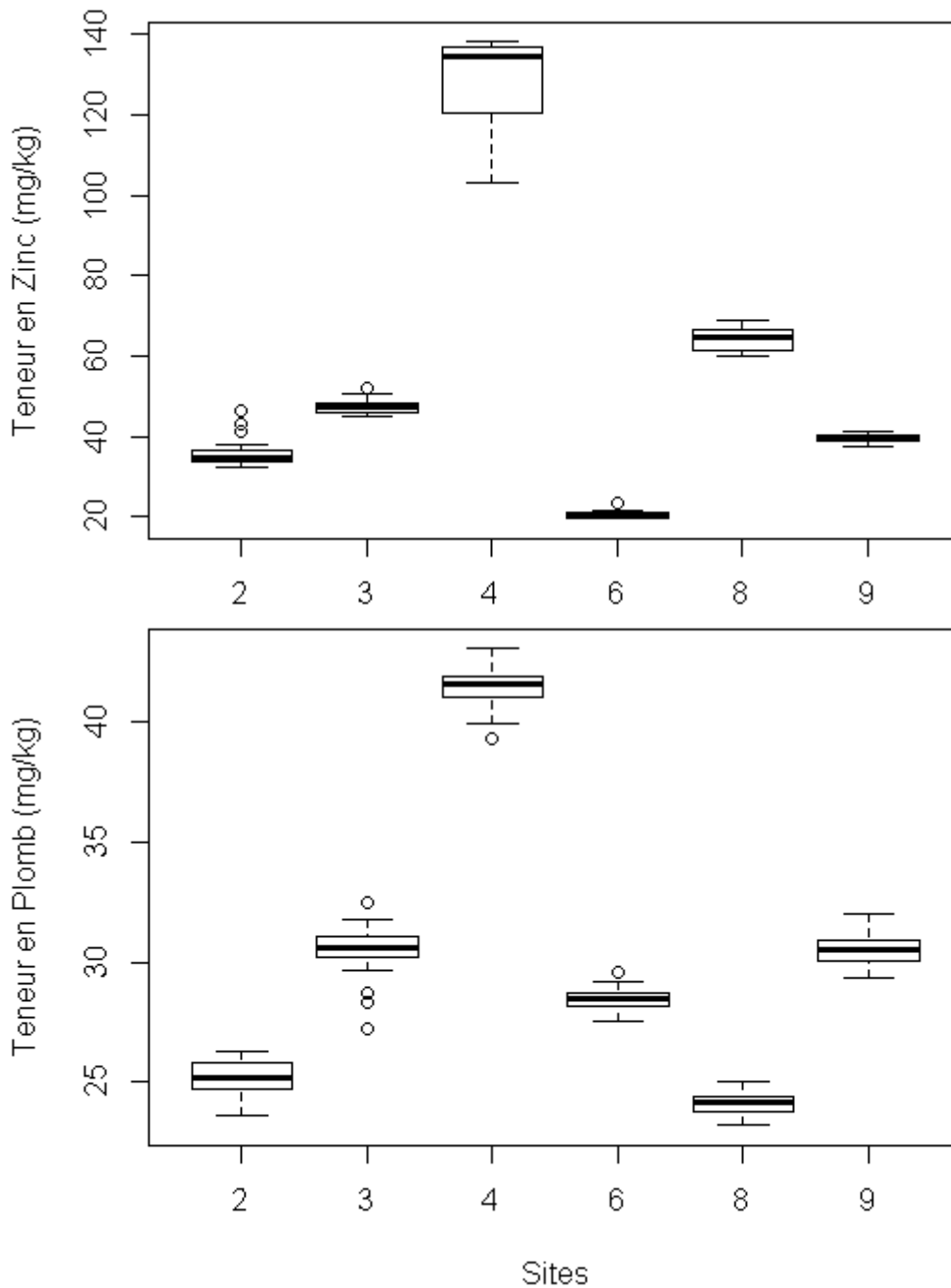


Figure 9 : Boxplots des teneurs en zinc et en plomb par site pour la couche de sub-surface

2.2.2. Comparaison des différents modèles de calcul de la variance des teneurs en éléments

2.2.2.1 Comparaison des modèles pour la couche de surface

Trois modèles mixtes hiérarchiques différents du calcul de la variance des teneurs en éléments sont ajustés pour chacune des variables puis comparés au moyen d'un test comprenant un test de ratio de vraisemblance et de comparaison d'AIC (*Akaike Information Criterion*) et d'une analyse des résidus. Le premier modèle testé n'inclut aucun effet fixe uniquement les effets aléatoires hiérarchisés, le deuxième modèle inclut un effet fixe lié à

l'occupation du sol et le troisième modèle intègre quant à lui un effet fixe lié au site (annexe 6). Ainsi, pour le molybdène les résultats du test privilégient un effet « site ». En effet l'AIC (*Akaike Information Criterion*) est le plus faible pour le troisième modèle. L'analyse des résidus pour le modèle sans effet fixe et avec effet fixe « site » (figure 10) confirme le résultat du test, les résidus semblent mieux se rapprocher d'une distribution normale dans le cas du modèle avec effet fixe « site ». Cependant, les résidus ne paraissent toujours pas normaux. De la même façon pour le nickel, le test tend à privilégier le troisième modèle mais l'analyse des résidus ne confirme pas tout à fait ce résultat (figure 11). En effet le modèle avec un effet fixe site ne change en rien la distribution des résidus. Au contraire des modèles du calcul de la variance pour le molybdène, ceux pour le nickel semblent meilleurs avec des résidus tout à fait proches d'une distribution normale. Ainsi, on gardera pour le molybdène un modèle avec un effet fixe lié au site en plus des effets aléatoires hiérarchisés et pour le nickel un modèle avec un effet fixe lié à l'occupation du sol.

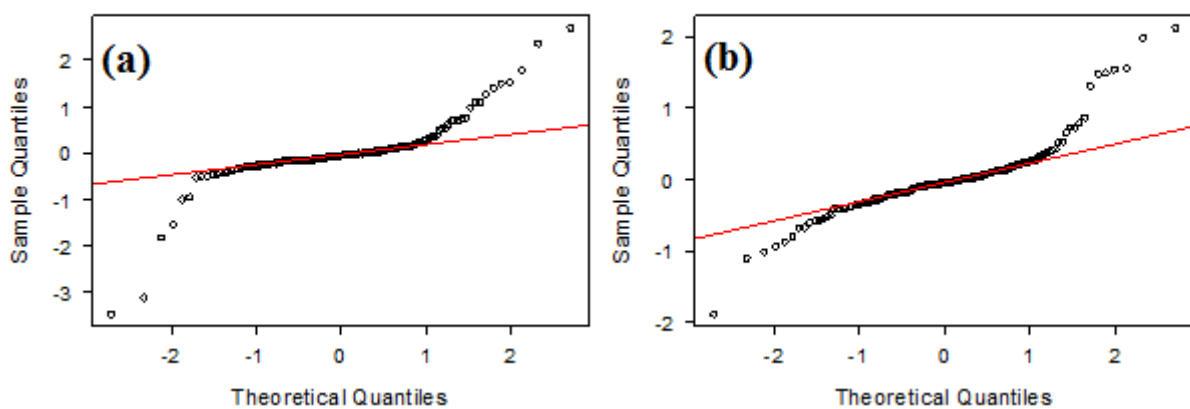


Figure 10: Diagrammes Quantile-Quantile des résidus pour les modèles sans effet fixe (a) et avec effet fixe site (b) du molybdène pour la couche de surface

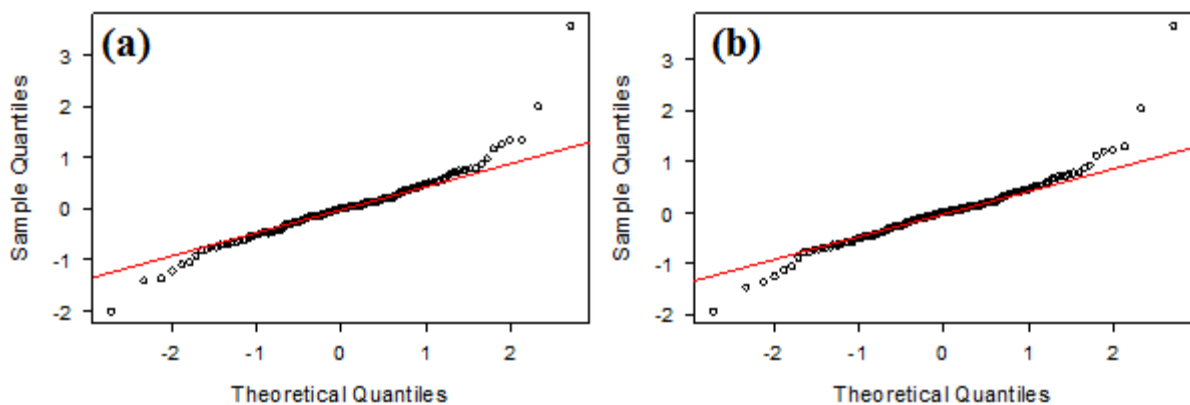


Figure 11 : Diagrammes Q-Q des résidus pour les modèles avec effet fixe occupation (a) et effet fixe site (b) du nickel pour la couche de surface

2.2.2.2 Comparaison des modèles pour la couche de sub-surface

Les résultats de la comparaison des modèles (annexe 6) désignent le modèle avec effet « site », pour le plomb et le zinc, comme étant le meilleur modèle parmi ceux testés pour le calcul de la variance des teneurs dans le sol. Cependant, pour les deux éléments l'analyse des résidus (figures 12 et 13) ne montre pratiquement aucune différence entre le modèle avec effet

fixe « site » et le modèle sans effet fixe uniquement constitué des effets aléatoires hiérarchisés. Ainsi, le modèle avec effet fixe « site » sera écarté en raison de sa complexité par rapport au modèle simple sans effet fixe. D'autre part, on remarquera que pour le cas du modèle de calcul de la variance de la teneur en plomb, les résidus semblent être très proches d'une distribution normale. De ce fait, on peut dire que le modèle sans effet fixe pour le plomb semble être bien ajusté. À l'inverse, les résidus correspondant au modèle pour le zinc sont très éloignés d'une loi normale ce qui indique que le modèle nécessite d'être ajusté.

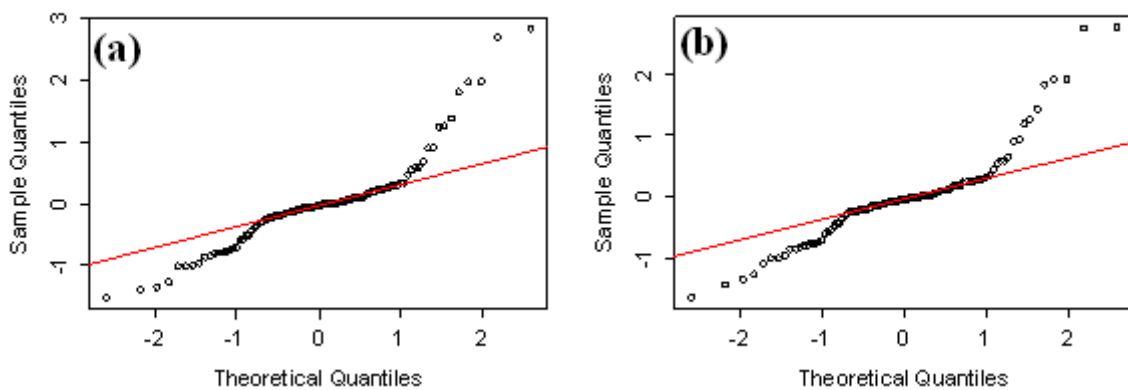


Figure 12 : Diagrammes Q-Q des résidus pour les modèles sans effet fixe (a) et effet fixe site (b) du zinc pour la couche de sub-surface

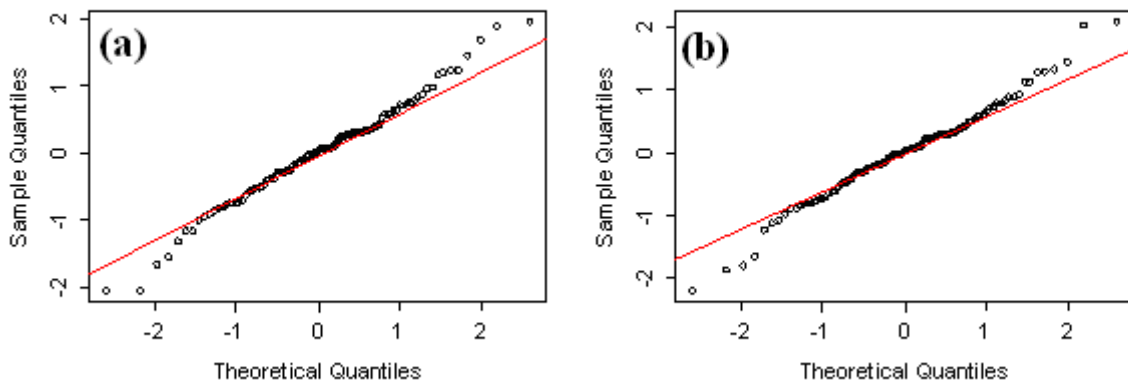


Figure 13 : Diagrammes Q-Q des résidus pour les modèles sans effet fixe (a) et effet fixe site (b) du plomb pour la couche de sub-surface.

2.2.3. Transformation des variables

2.2.3.1. Transformation logarithmique

Une première transformation naturelle consiste au passage au logarithme pour la variable étudiée dans le modèle. Pour certains éléments, le passage au logarithme ne permet pas d'ajuster au mieux le modèle. Cependant, il existe des éléments pour lesquels cette transformation semble améliorer le modèle. C'est le cas par exemple du modèle de calcul de la variance du cuivre et du nickel. L'observation des résidus des modèles associés au calcul de la variance de ces deux éléments traces (figures 14 et 15) montre en effet un meilleur ajustement des valeurs extrêmes.

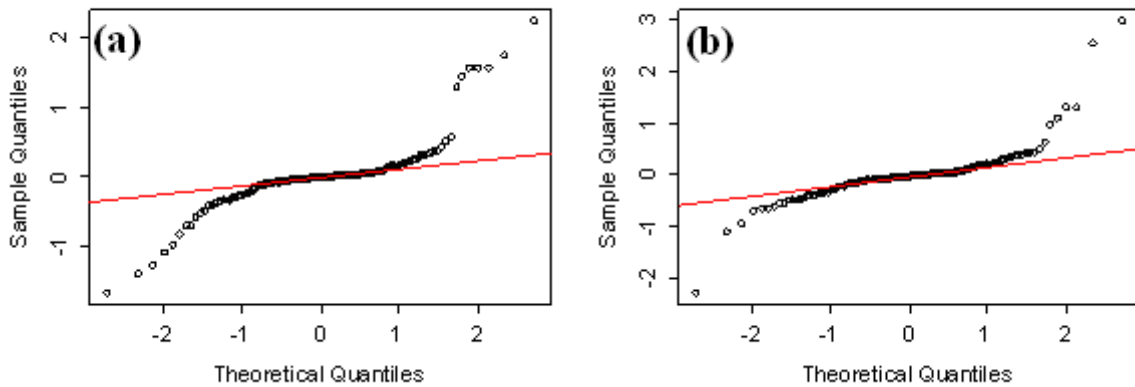


Figure 14 : Q-Q plot des résidus pour le modèle avec effet occupation sans (a) et avec (b) transformation logarithmique du cuivre pour la couche de surface.

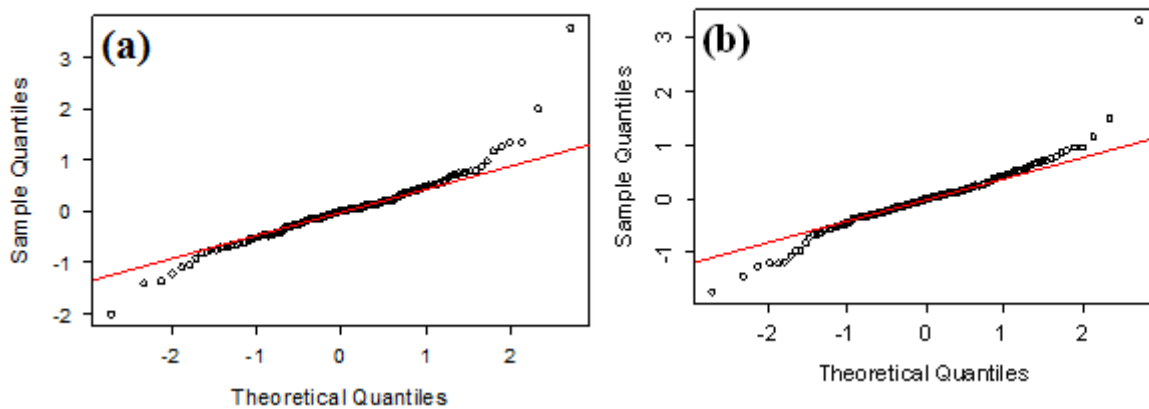


Figure 15 : Q-Q plot des résidus pour le modèle avec effet occupation sans (a) et avec (b) transformation logarithmique du nickel pour la couche de surface.

2.2.3.2. Transformation de Box-Cox

Les analyses de résidus précédentes sur les variables étudiées montrent une distribution des résidus éloigné d'une loi normale. En particulier, les résidus issus du modèle de calcul de la variance des teneurs en nickel pour la couche de surface et du cobalt pour la couche de sub-surface ne sont pas normaux. Ainsi, on transforme cette fois la variable de départ à l'aide d'une transformation de Box-Cox.

Pour quelques éléments chimiques étudiés, la transformée de Box Cox améliorent le modèle. En effet, pour le nickel, le cadmium, le cuivre et le cobalt pour la première couche de sol ainsi que pour le cobalt et le zinc pour la couche de sub-surface la transformation de Box-Cox corrige de façon plus ou moins forte selon les variables les résidus issus du modèle et les rapprochent d'une distribution normale.

Ainsi, les figures 16 et 17 représentent les diagrammes quantile-quantile respectivement des modèles de calcul de la variance des teneurs en nickel pour la couche de surface et des teneurs en cobalt pour la couche de sub-surface avant et après transformation. On remarque que pour le nickel, comme pour la transformation logarithmique, la transformation ici opérée permet d'ajuster au mieux les valeurs extrêmes et la distribution des résidus du modèle semblent tout à fait normale. Pour le cobalt, alors que les résidus du modèle de calcul de la variance étaient loin de suivre une loi gaussienne, la transformation de Box-Cox a permis ici de grandement les rapprocher d'une distribution normale.

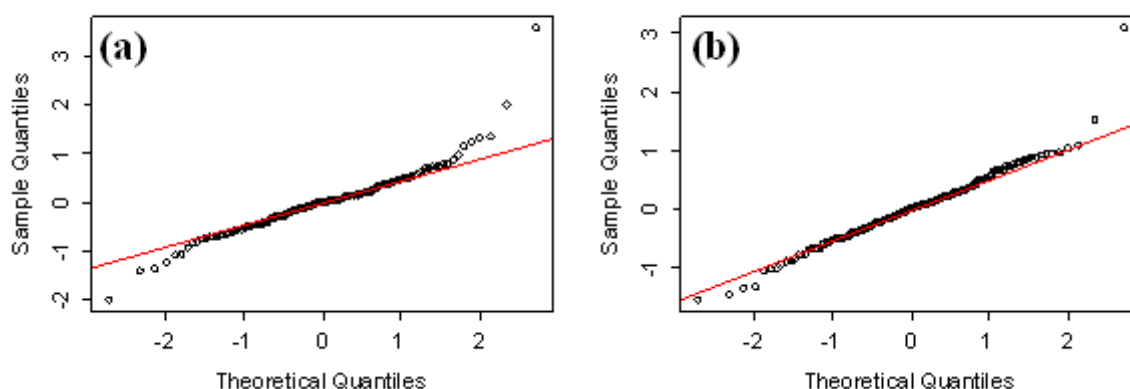


Figure 16 : Q-Q plot des résidus pour le modèle avec effet fixe occupation sans (a) et avec (b) transformation de Box-Cox du nickel pour la couche de surface

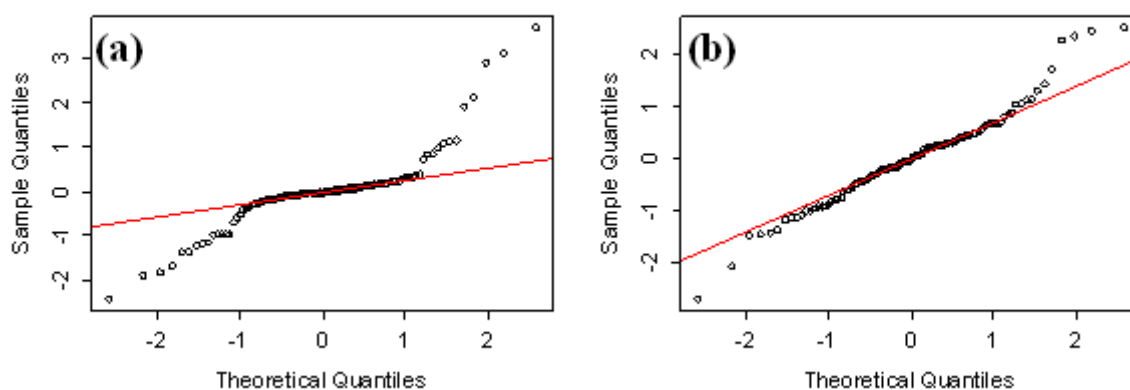


Figure 17 : Q-Q plot des résidus pour le modèle sans effet fixe, sans (a) et avec (b) transformation de Box-Cox du cobalt pour la couche de sub-surface.

2.3. Composants de la variance

Pour chacune des variables étudiées, on présente les composants de la variance estimés pour la couche de surface et de sub-surface issus des modèles linéaires mixtes hiérarchiques qui minimisent l'AIC (*Akaike Information Criterion*) (Annexes 9 et 10). Le composant de la variance d'une étape donnée est la variance de la population pour l'effet aléatoire associé. Les tableaux 2 et 3 présentent les composants de la variance pour les variables sélectionnées.

Ainsi, pour les deux couches de sol, le composant de la variance correspondant à l'effet aléatoire « site » est le plus élevé pour toutes les variables. Le composant de la variance le plus faible ne correspond jamais aux mêmes effets aléatoires d'une variable à l'autre. Cependant, on note que la plupart des variances à l'exception des variances associées à l'effet site pour le nickel et pour la première couche de sol ont tendance à être très petites et à tendre vers zéro. D'autre part, on observe en annexe que pour toutes les variables les composants de la variance sont plus faibles pour la couche de sub-surface.

Tableau 2 : Composants de la variance estimés du modèle qui minimisent l'AIC pour le Nickel et le Molybdène pour la couche de surface

Eléments	σ_{site}	σ_{comp}	σ_{id2}	σ_{id3}	σ_{res}
Molybdène	5.9×10^{-1}	3.3×10^{-2}	1.9×10^{-2}	3.9×10^{-2}	3.1×10^{-2}
Nickel	1.1×10^1	3.3×10^{-1}	2.9×10^{-1}	4.3×10^{-1}	3.2×10^{-1}

Tableau 3 : Composants de la variance estimés du modèle qui minimisent l'AIC pour Plomb et le Zinc pour la couche de sub-surface

Eléments	σ_{site}	σ_{comp}	σ_{id2}	σ_{id3}	σ_{res}
Plomb	6.2×10^{-3}	7.9×10^{-4}	3.2×10^{-4}	4.3×10^{-4}	5.2×10^{-4}
Zinc	3.6×10^{-2}	1.6×10^{-3}	4.3×10^{-3}	1.9×10^{-3}	2.2×10^{-3}

Conclusion

Un schéma d'échantillonnage et de préparation des échantillons est mis en place dans le cadre du Réseau de Mesures de la Qualité des Sols. L'objectif était d'évaluer les incertitudes liées aux différentes étapes analytiques de préparation des échantillons pour les mesures de teneurs de plusieurs contaminants dans le sol. Les variances de chacune des étapes analytiques des teneurs de dix éléments chimiques différents (N, Cd, Cr, Co, Cu, Mo, Ni, Pb, Tl, Zn) sont ainsi calculées. Dans cette étude, nous avons également cherché à vérifier l'efficacité du schéma déséquilibré d'échantillonnage du RMQS établie.

Ainsi, des modèles linéaires mixtes hiérarchiques sont mis en œuvre à partir du jeu de données issu du schéma déséquilibré du RMQS. Le but de ces modèles est alors d'estimer les variances des teneurs des différents éléments. Nous avons ajusté ces modèles mixtes hiérarchiques en cherchant tout d'abord des variables pertinentes à incorporer aux modèles. Les modèles sont ensuite comparés, pour chacune des variables, au moyen d'un test de ratio de vraisemblance et de comparaison des AIC afin de ne retenir que le meilleur. Pour ajuster le modèle, on procède également à des transformations sur les variables de départ. L'efficacité du schéma d'échantillonnage est vérifiée par simulation de jeux de données sur lesquels on estime les variances par recherche du maximum de vraisemblance.

Les résultats ainsi obtenus à la suite des simulations de jeux de données associés au modèle mixte hiérarchique pour différents schémas nous a permis de montrer l'efficacité du schéma de mesure de répétabilité établie par le RMQS. En effet, couplé aux méthodes d'estimation par REML, le schéma du RMQS permet d'estimer les variances liées aux teneurs des contaminants avec peu d'incertitudes. Les résultats des simulations montrent ainsi l'absence de biais dans l'estimation des paramètres pour la structure du schéma RMQS. Cependant, la précision reste faible ne permettant pas de distinguer statistiquement les différentes sources d'erreurs. L'ajustement de modèle mixte hiérarchique aboutit à l'incorporation d'effets fixes différents pour chacune des variables. En effet, alors qu'un effet « site » est privilégiée pour des variables, un effet occupation voire aucun effet fixe est retenue pour d'autres variables. Ainsi, il est impossible d'ajuster un unique modèle linéaire mixte hiérarchique pour l'ensemble des éléments chimiques ici étudié. De plus, les transformations logarithmiques et de Box-Cox s'avèrent efficace pour quelques variables seulement. De la même façon l'application d'une transformation unique efficace pour l'intégralité des variables n'est pas possible. Enfin, l'application de ce modèle aux données du RMQS montre la difficulté à trouver un modèle respectant les hypothèses statistiques

Il est possible par la suite d'utiliser les résultats d'estimation des variances des teneurs en contaminants pour l'appliquer au calcul de l'erreur associée au stock de chacun des contaminants.

Bibliographie

Corsten L. C. A. and Stein A., 1994, *Nested sampling for estimating spatial semivariograms compared to other designs*, Applied stochastic models and data analysis, Vol.10, 103-122.

Goidts E., Van Wesemael B., Crucifix M., 2009, *Magnitude and sources of uncertainties in soil organic carbon (SOC) stock assessments at various scales*, European Journal of Soil Science, 60, 723-739.

Jolivet C., Arrouays D., Boulonne L., Ratié C., Saby N., 2006, *Le Réseau de Mesures de la Qualité des sols de France (RMQS) : Etat d'avancement et premiers résultats*, Etude et Gestion des Sols, 13, 149-164.

Jolivet C., Boulonne L., Ratié C., 2006, *Manuel du Réseau de Mesures de la Qualité des Sols de France*, édition 2006, Unité InfoSol, INRA Orléans, France, 190p.

Lark R.M., 2011, *Spatially nested sampling schemes for spatial variance components: Scope for their optimization*, Computers & Geosciences, doi:10.1016/j.cageo.2010.12.010

Papritz A., Dumig A., Zimmermann C., Gerke H. H., Felderer B., Kögel-Knabner I., Schaaf W. & Schulín R., 2011, *Uncertainty of variance component estimates in nested sampling: a case study on the field-scale spatial variability of a restored soil*, European Journal of Soil Science, 62, 479-495.

Pinheiro J.C., and Bates D.M., 2000, *Mixed-Effects Models in S and S-PLUS*, Springer, 528p.

Rawlins B.G., Scheib A. J., Lark R.M., Lister T.R., *Sampling and analytical plus subsampling variance components for five soil indicators observed at regional scale*, European Journal of Soil Science, 60, 740-747.

Webster R., Welham S.J., Potts J.M., Oliver M.A., 2006, *Estimating the spatial scales of regionalized variables by nested sampling, hierarchical analysis of variance and residual maximum likelihood*, Computer & Geosciences, 32: 1320-1333.

Annexe 1 : Statistiques descriptives des sites de répétabilité pour la couche de surface

Teneur en éléments traces (mg/kg) et en azote (g/kg)

	Moyenne	Var	Min	Max		Moyenne	Var	Min	Max
Site 269					Site 1208				
N	2,69	0,0025	2,61	2,78		1,60	0,0086	1,42	1,79
Cd	0,34	0,00020	0,32	0,37		0,59	0,023	0,50	1,22
Cr	49,63	1,05	47,80	51,10		48,53	7,41	39,60	52,64
Co	9,49	0,044	9,10	9,83		21,26	2,29	18,90	24,50
Cu	19,08	0,11	18,80	20,10		10,20	1,38	9,18	14,24
Mo	0,56	0,00055	0,54	0,63		0,72	0,0013	0,68	0,84
Ni	23,24	0,20	22,6	24,1		24,89	0,73	23,11	26,28
Pb	30,52	1,52	29,5	33,9		34,53	0,91	32,14	35,74
Tl	0,54	1,77E-05	0,54	0,55		0,75	0,00061	0,70	0,80
Zn	81,86	11,81	78,10	90,90		125,31	17,29	118,89	131,19
Site 689					Site 1326				
N	3,31	0,013	3,13	3,52		3,92	0,019	3,75	4,15
Cd	0,35	0,00076	0,32	0,42		0,84	0,00044	0,81	0,88
Cr	34,14	25,35	24,30	39,50		74,12	3,23	72,00	77,10
Co	5,18	0,032	4,81	5,48		9,17	0,042	8,86	9,49
Cu	11,07	0,68	9,65	12,30		16,55	0,082	16,20	17,00
Mo	0,30	0,0021	0,27	0,48		1,58	0,00041	1,55	1,61
Ni	14,15	0,19	13,5	14,85		37,23	0,97	36	38,8
Pb	28,62	2,10	27,02	33		36,98	0,63	36	38
Tl	0,56	0,0002	0,54	0,59		1,12	0,00063	1,08	1,15
Zn	52,58	5,35	49,78	56,90		148,50	13,91	143,00	156,00
Site 801					Site 10797				
N	1,88	0,0074	1,78	2,12		0,91	0,0095	0,82	1,10
Cd	0,27	0,00064	0,25	0,37		0,048	5,63E-05	0,039	0,063
Cr	29,88	13,45	21,60	34,60		15,20	1,51	12,90	17,20
Co	6,13	0,011	5,97	6,36		3,44	0,035	3,16	3,70
Cu	6,98	0,044	6,59	7,45		2,88	0,019	2,64	3,03
Mo	0,27	0,0010	0,25	0,39		0,26	0,00027	0,23	0,28
Ni	14,01	0,094	13,5	14,7		5,47	0,11	4,96	6,05
Pb	28,69	0,38	27,5	30,2		35,03	10,20	32	43,8
Tl	0,79	0,00033	0,74	0,82		0,86	0,00088	0,81	0,91
Zn	36,32	6,47	34,07	46,10		19,60	0,81	17,80	21,10

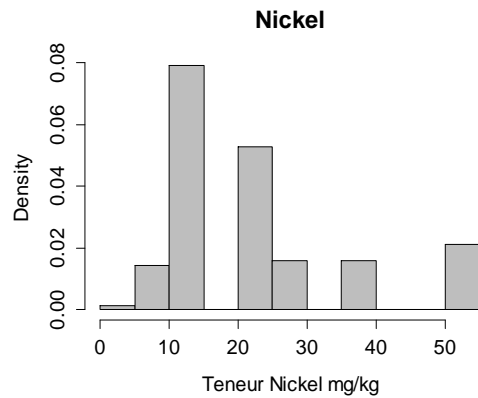
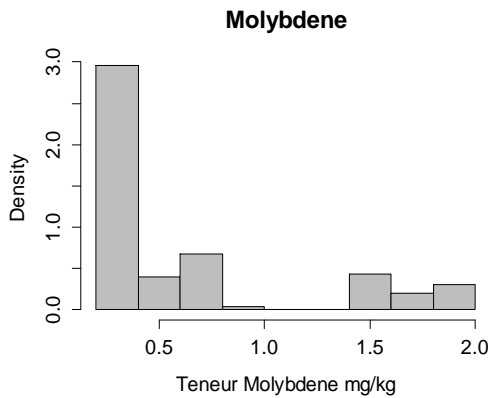
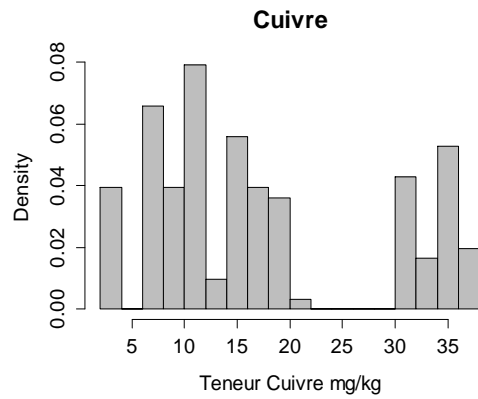
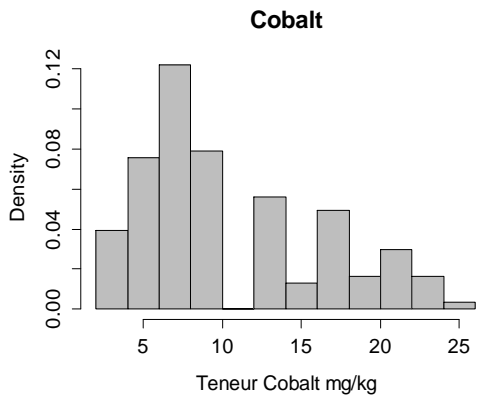
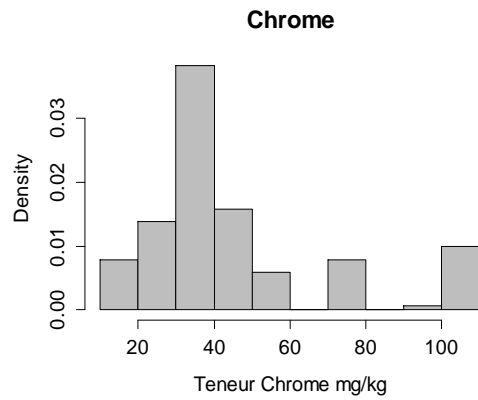
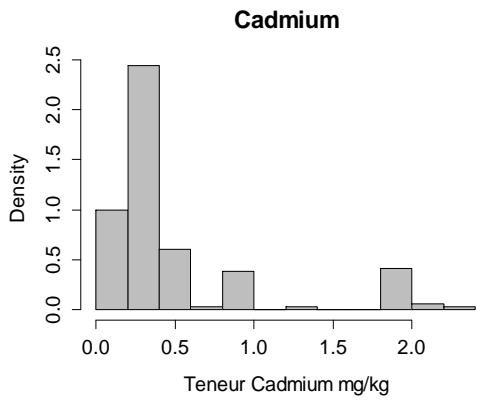
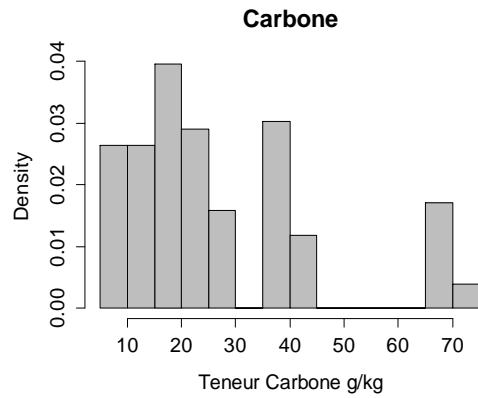
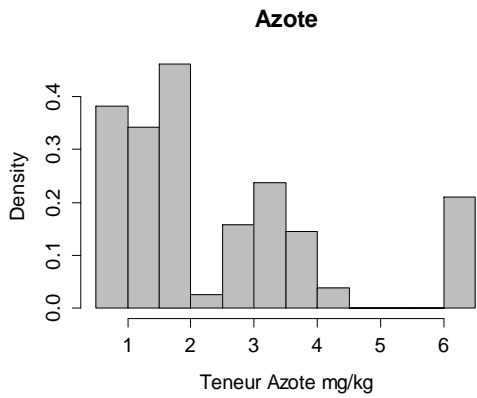
Site 899					Site 11151				
N	1,36	0,0014	1,30	1,43	6,30	0,0025	6,21	6,38	
Cd	0,28	7,96E-05	0,26	0,30	2,00	0,0074	1,89	2,23	
Cr	31,82	30,95	20,70	37,05	102,96	3,11	99,30	107	
Co	13,47	0,39	12,69	15,21	16,31	0,053	15,60	16,70	
Cu	35,68	0,35	34,75	37,05	14,94	0,15	14,30	15,60	
Mo	0,28	0,00068	0,26	0,34	1,78	0,018	1,51	1,93	
Ni	23,84	0,36	23,2	25,19	53,13	0,61	51	54,4	
Pb	37,04	5,35	34,9	46,3	47,79	0,89	46,5	49,5	
Tl	1,028	0,0032	0,90	1,1	0,79	0,0093	0,64	1,00	
Zn	50,57	4,16	48,10	55,38	280,81	22,83	269	288	
Site 989									
N	0,72	0,0023	0,68	0,89					
Cd	0,19	5,8E-05	0,18	0,21					
Cr	36,12	14,61	27,99	41,30					
Co	6,25	0,022	6,05	6,54					
Cu	32,09	3,36	30,30	37,90					
Mo	0,30	0,00011	0,27	0,32					
Ni	14,23	0,088	13,8	14,80					
Pb	30,73	0,90	29,8	33,42					
Tl	1,05	0,0022	1,002	1,22					
Zn	50,87	47,63	46,40	79,30					

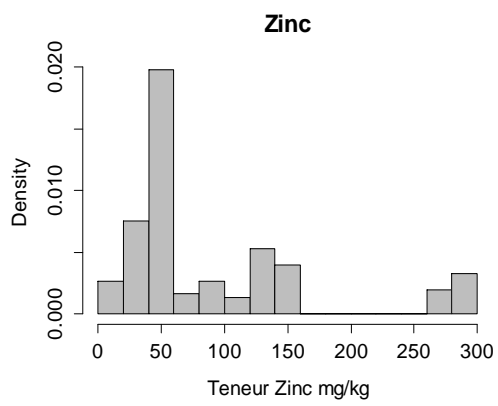
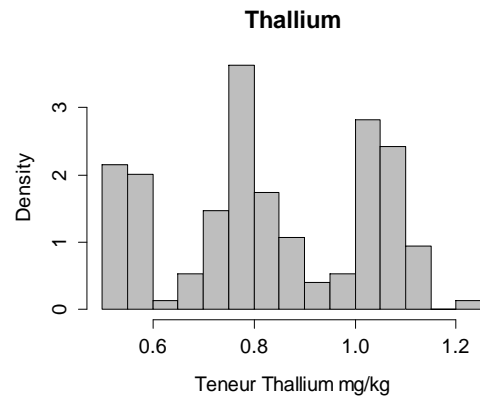
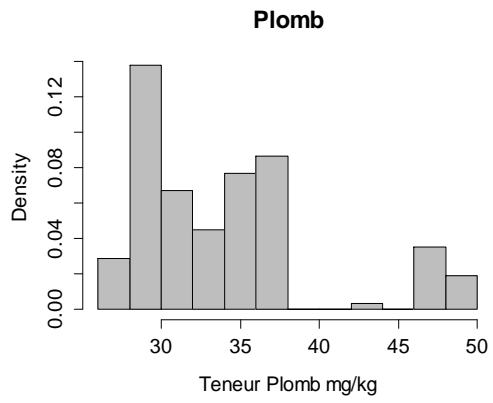
Annexe 2 : Statistiques descriptives des sites de répétabilité pour la couche de sub-surface

Teneur en éléments traces (mg/kg) et en azote (g/kg)

	Moyenne	Var	Min	Max		Moyenne	Var	Min	Max
Site 269					Site 989				
N	1,49	0,0038	1,40	1,58		0,33	0,00033	0,31	0,40
Cd	0,23	9,42E-05	0,21	0,25		0,13	4,18E-05	0,12	0,14
Cr	49,73	3,28	46,70	52,20		31,31	23,36	19,00	37,44
Co	10,10	0,11	9,68	10,80		6,59	0,025	6,35	6,89
Cu	13,23	0,15	12,70	14,00		9,07	0,67	8,39	11,70
Mo	0,51	7,82E-05	0,49	0,53		0,27	0,00040	0,25	0,34
Ni	23,99	0,42	23,10	25,20		15,40	0,18	14,70	16,29
Pb	24,12	0,23	23,20	25,00		30,50	0,46	29,30	32,02
Tl	0,56	3,65E-05	0,55	0,57		1,09	0,00030	1,05	1,12
Zn	64,29	8,02	60,10	68,70		39,37	0,91	37,30	41,40
Site 801					Site 1208				
N	0,92	0,0032	0,75	1,05		0,31	0,00026	0,26	0,34
Cd	0,22	6,13E-05	0,21	0,24		0,18	0,00057	0,15	0,25
Cr	29,70	49,33	18,70	35,68		59,77	30,70	47,50	69,90
Co	6,39	0,031	6,05	6,70		17,36	14,35	13,80	30,80
Cu	6,33	0,039	5,94	6,67		14,00	0,87	11,90	15,94
Mo	0,30	0,0059	0,24	0,54		0,98	0,0082	0,81	1,28
Ni	14,85	0,18	14,00	15,70		27,57	4,88	22,10	29,50
Pb	25,23	0,45	23,60	26,26		41,37	0,74	39,30	43,00
Tl	0,81	0,00072	0,74	0,85		1,06	0,00045	1,00	1,09
Zn	35,95	13,61	32,40	46,55		129,30	112,60	103,00	138,17
Site 899					Site 10797				
N	0,56	0,0014	0,52	0,63		0,39	0,0014	0,34	0,45
Cd	0,12	9,93E-05	0,11	0,14		0,04	1,05E-05	0,04	0,05
Cr	38,62	36,80	28,20	48,78		15,84	0,49	14,70	16,80
Co	13,96	1,50	11,80	16,20		4,10	0,0062	3,93	4,22
Cu	10,51	0,28	9,81	11,50		2,83	0,033	2,70	3,39
Mo	0,27	0,00041	0,24	0,34		0,22	9,63E-05	0,21	0,24
Ni	30,50	1,76	27,80	33,30		6,20	0,04	5,90	6,52
Pb	30,44	1,49	27,20	32,50		28,48	0,32	27,50	29,60
Tl	1,10	0,0043	0,95	1,18		0,91	0,00042	0,88	0,94
Zn	47,33	3,32	44,80	52,21		20,63	1,40	19,40	23,60

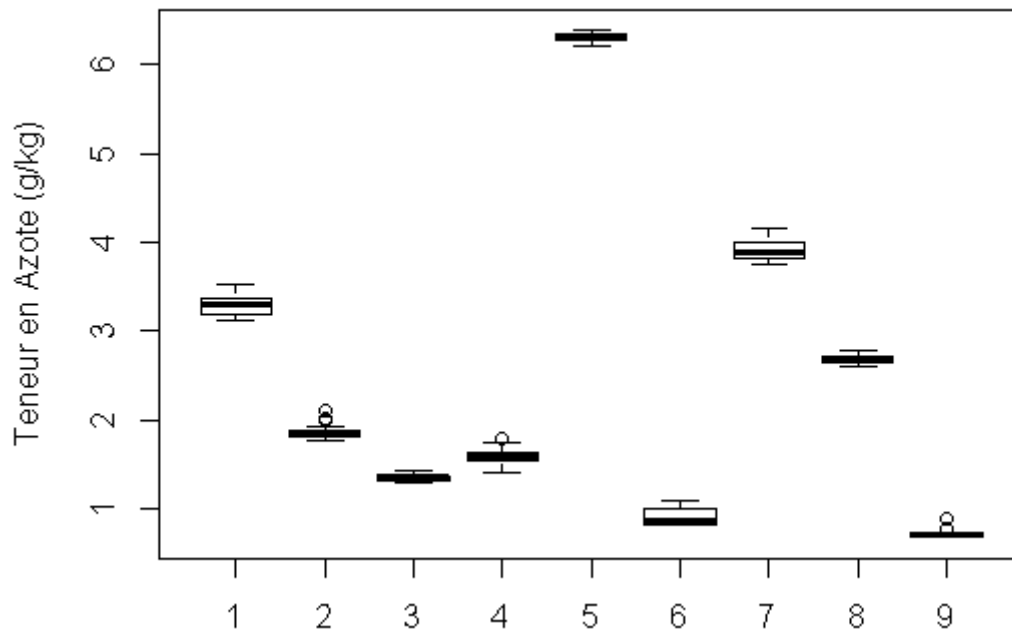
Annexe3 : Histogrammes de tous les éléments pour la couche de surface



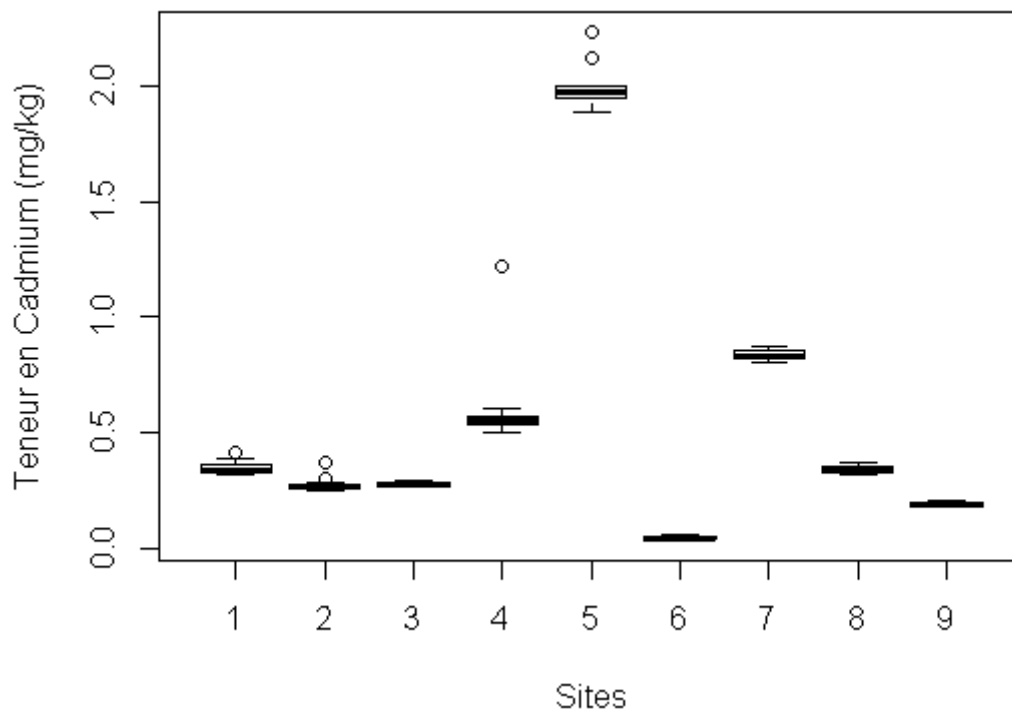


Annexe 4 : Boxplots des teneurs en éléments par site de répétabilité pour la couche de surface

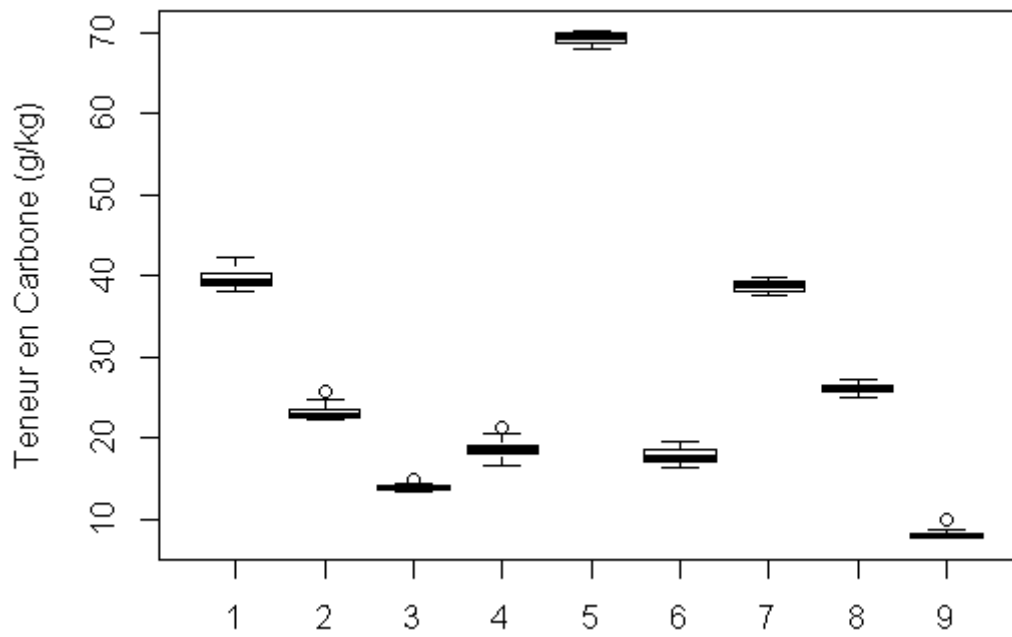
(a) Azote



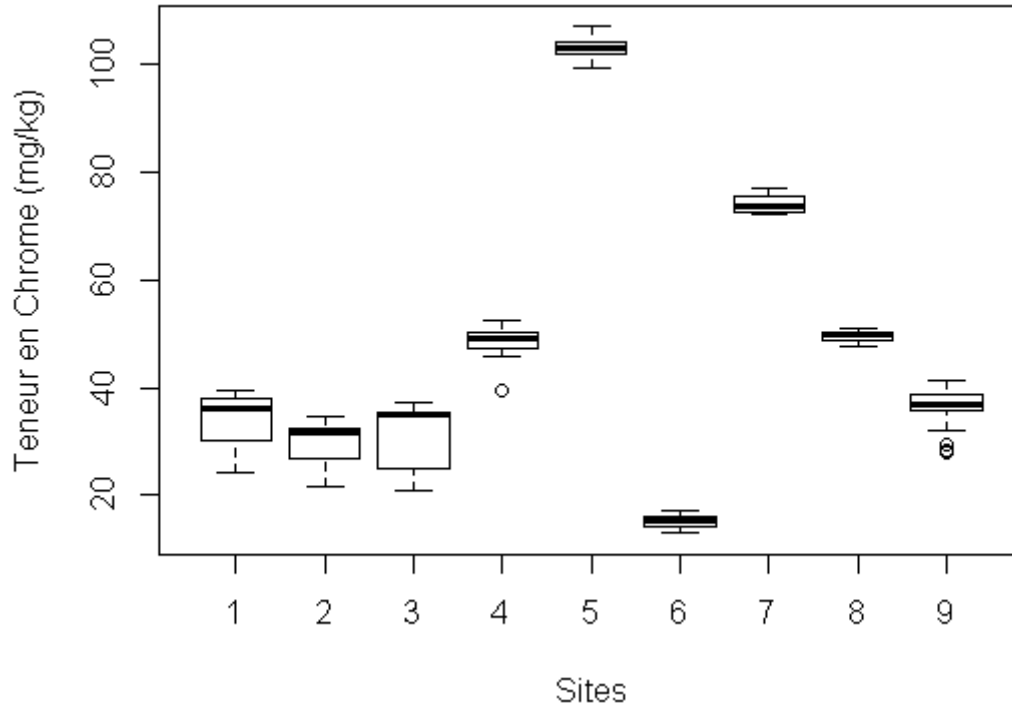
(b) Cadmium



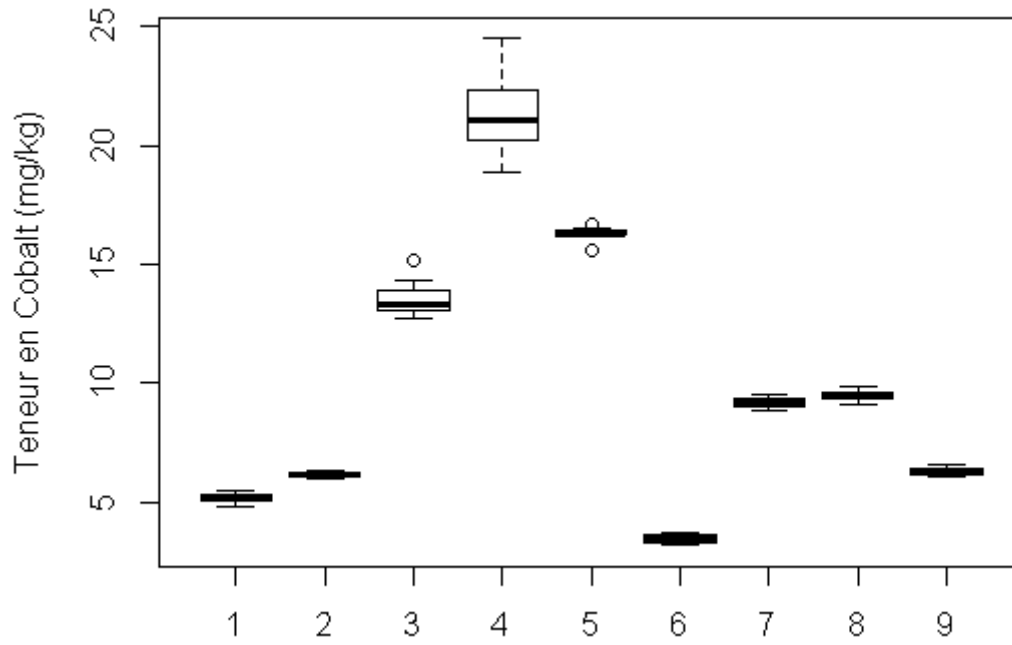
(c) Carbone



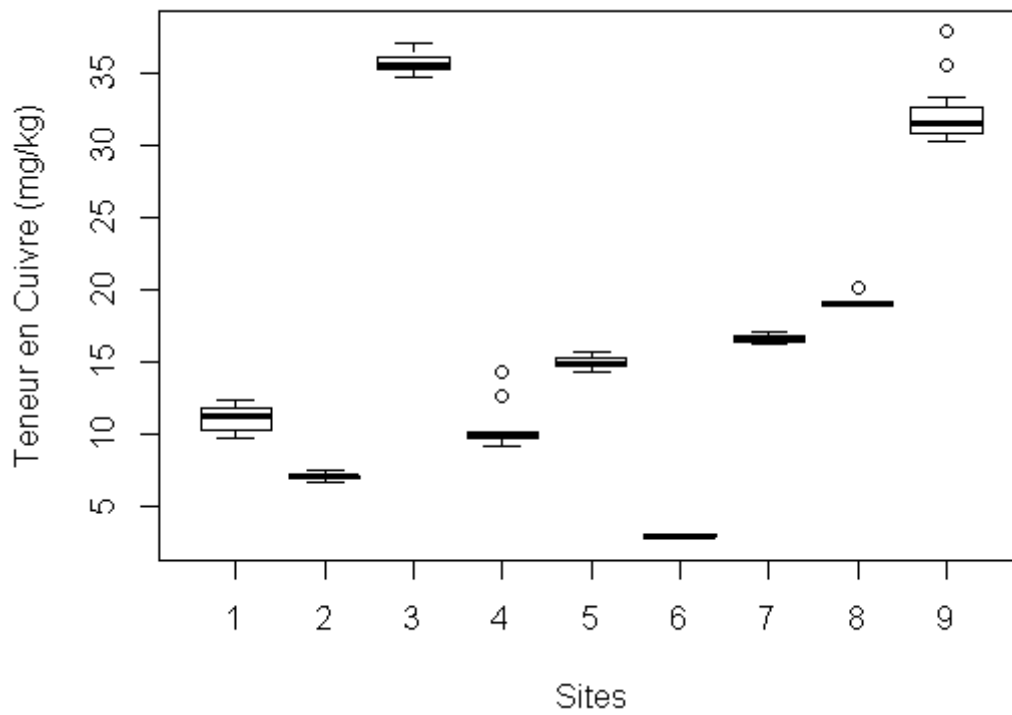
(d) Chrome



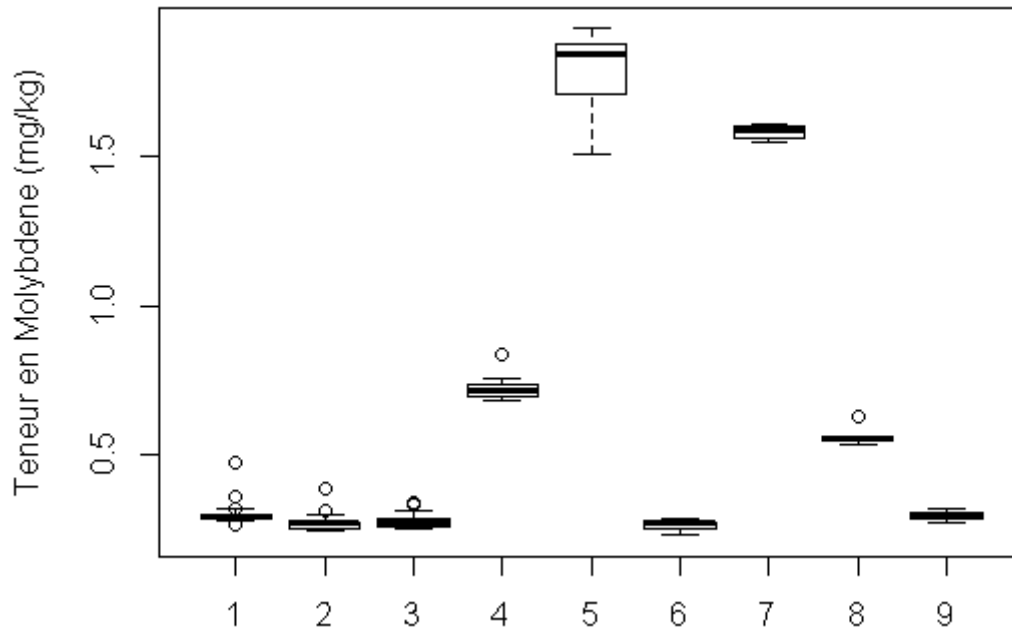
(e) Cobalt



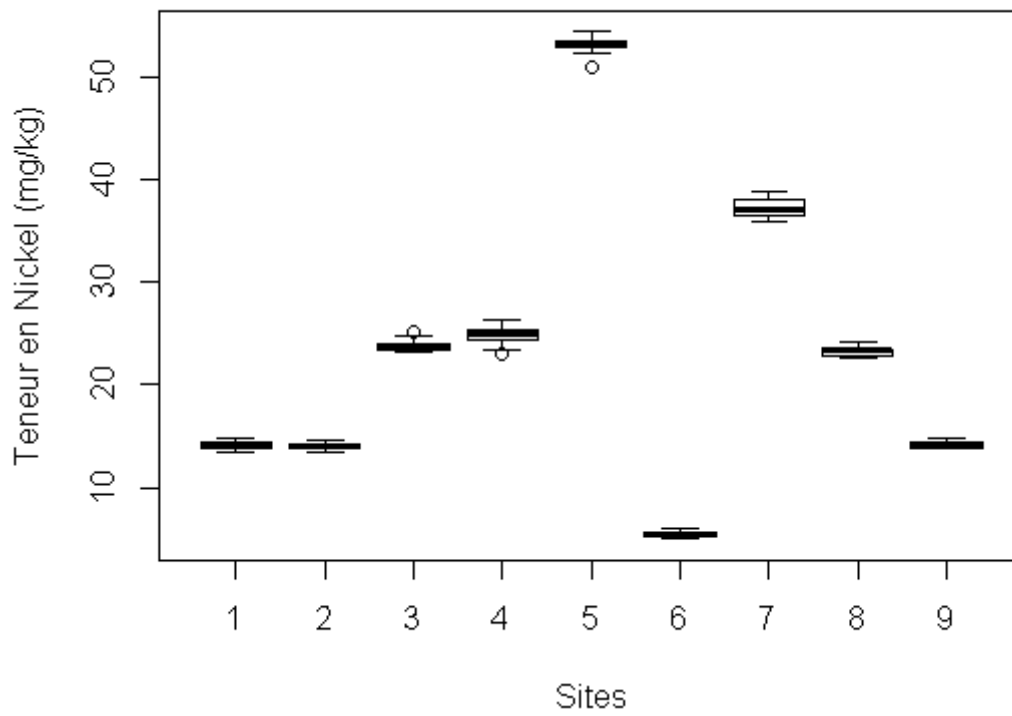
(f) Cuivre



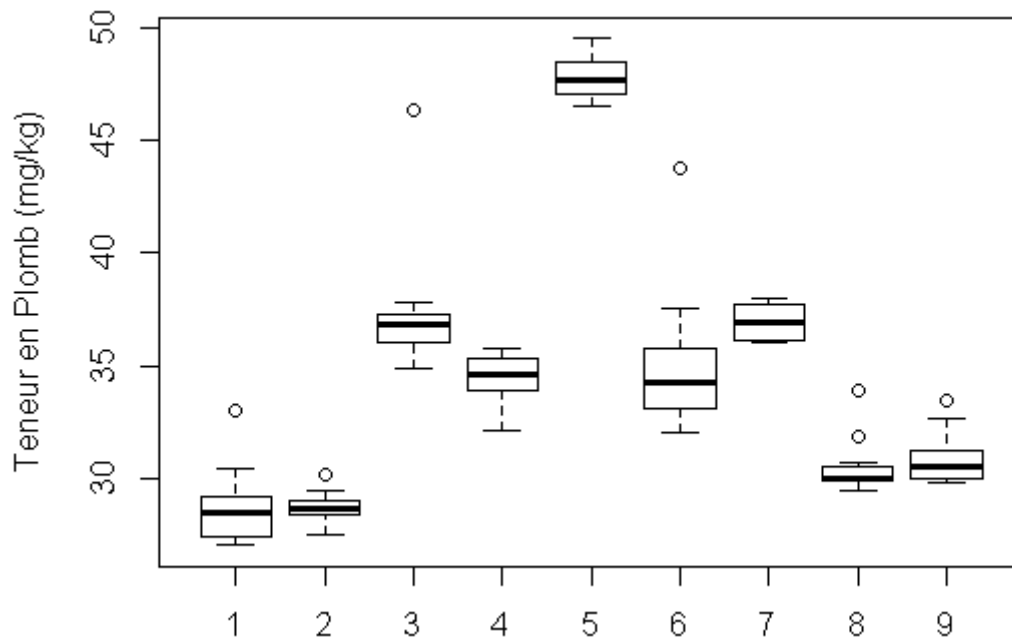
(g) Molybdène



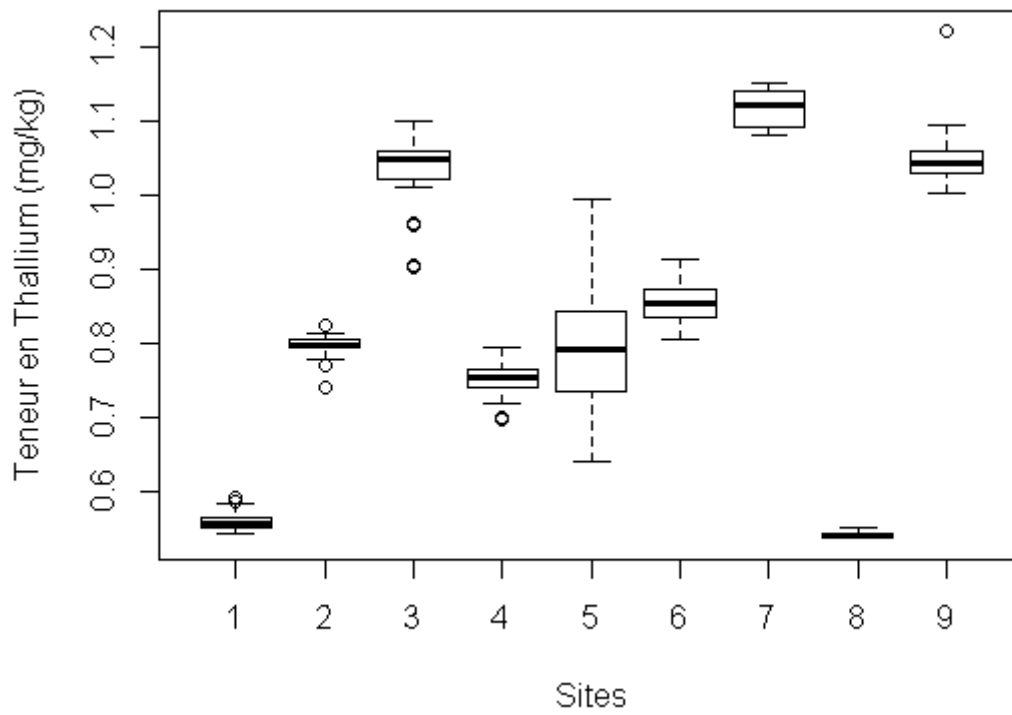
(h) Nickel



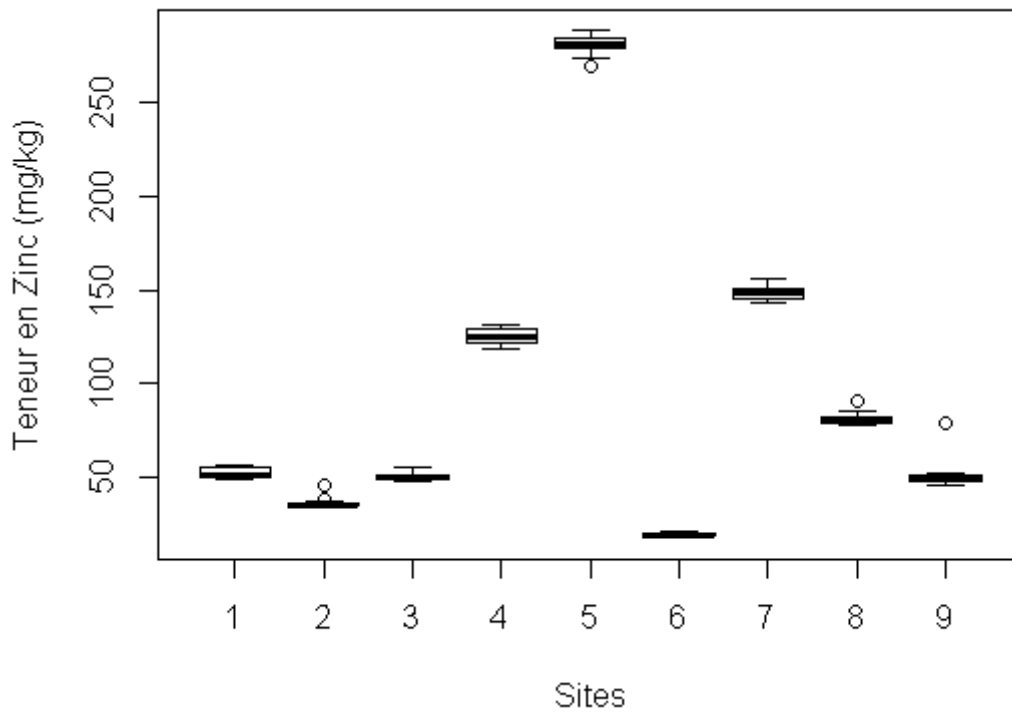
(i) *Plomb*



(j) *Thallium*

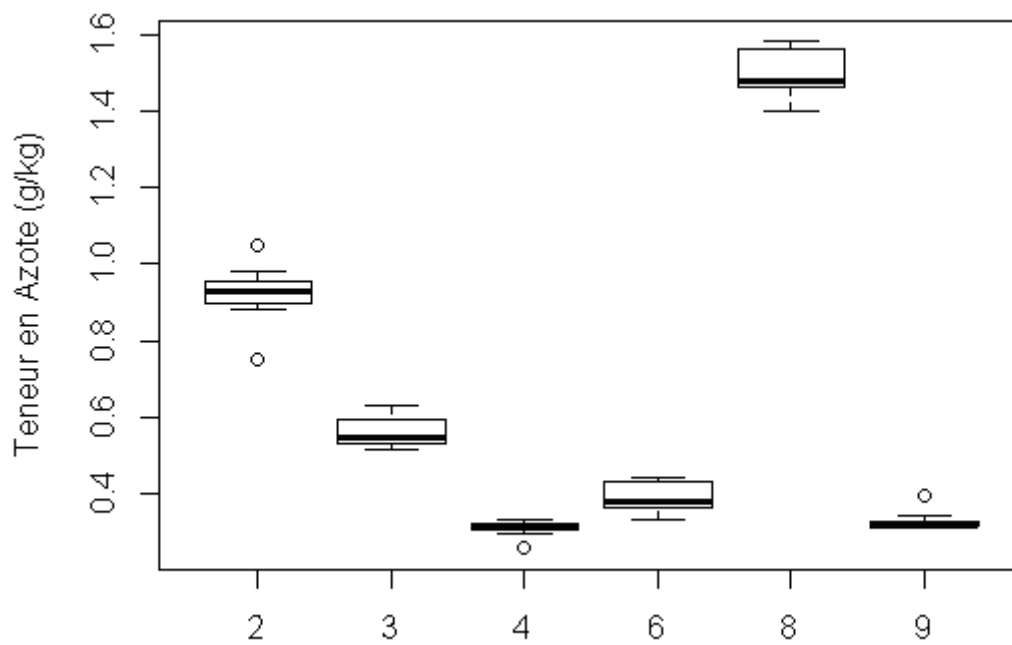


(k) Zinc

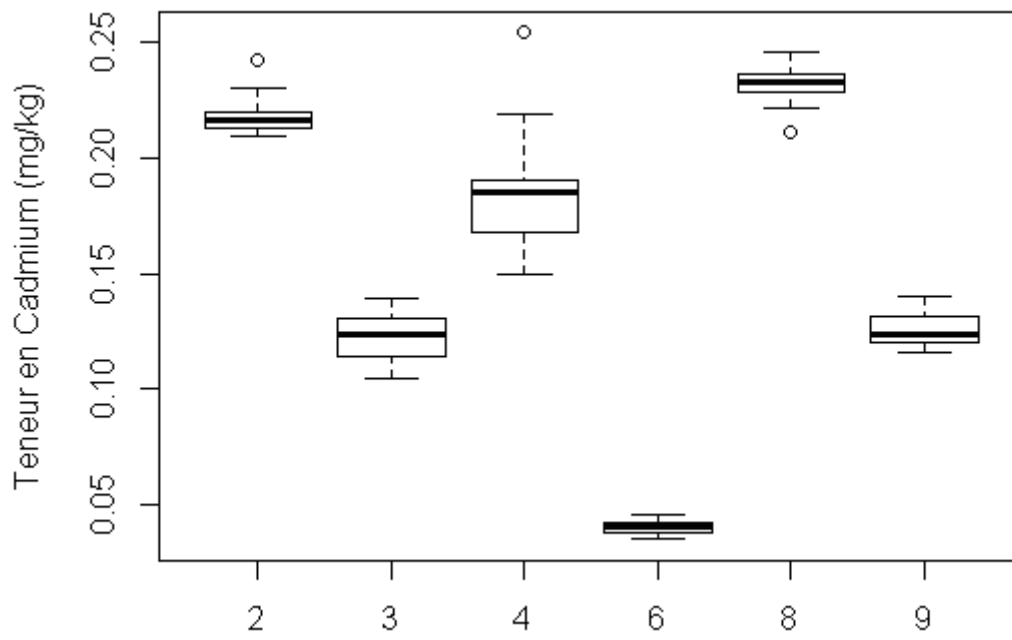


Annexe 5 : Boxplots des teneurs en éléments par site de répétibilité pour la couche de sub-surface

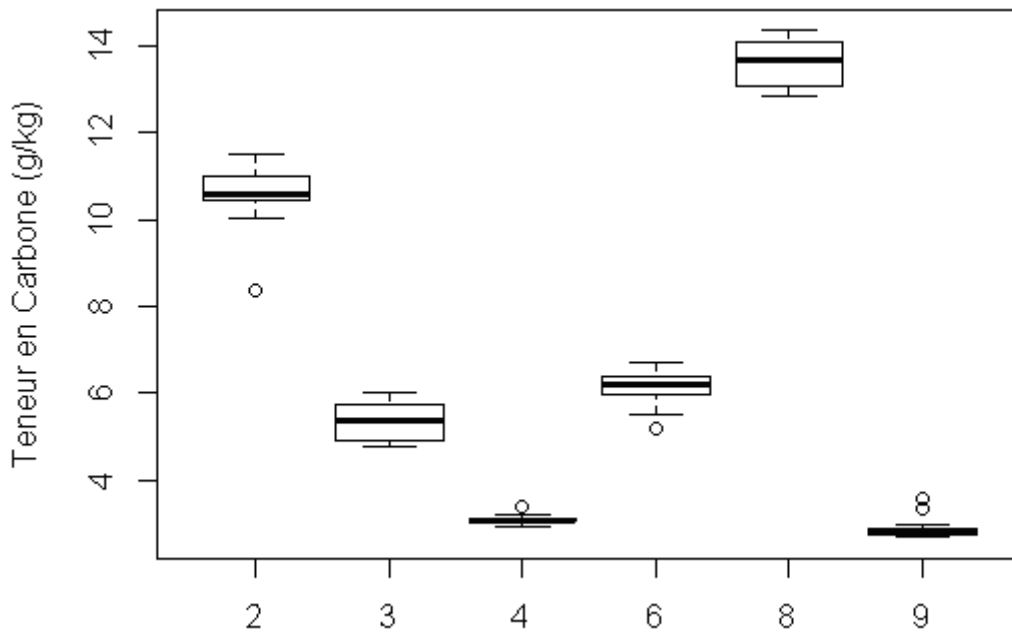
(a) Azote



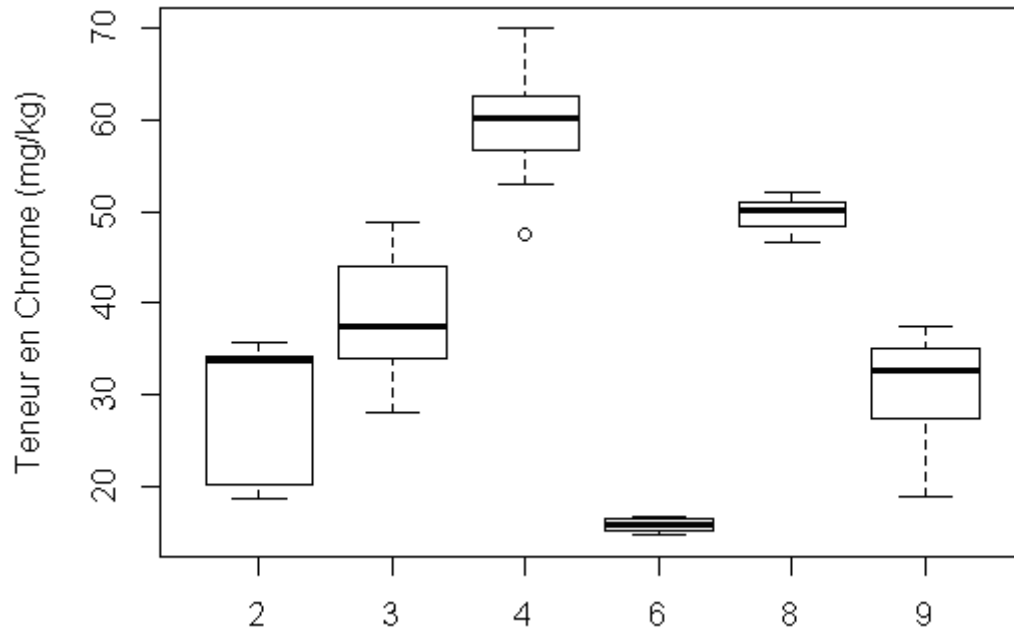
(b) Cadmium



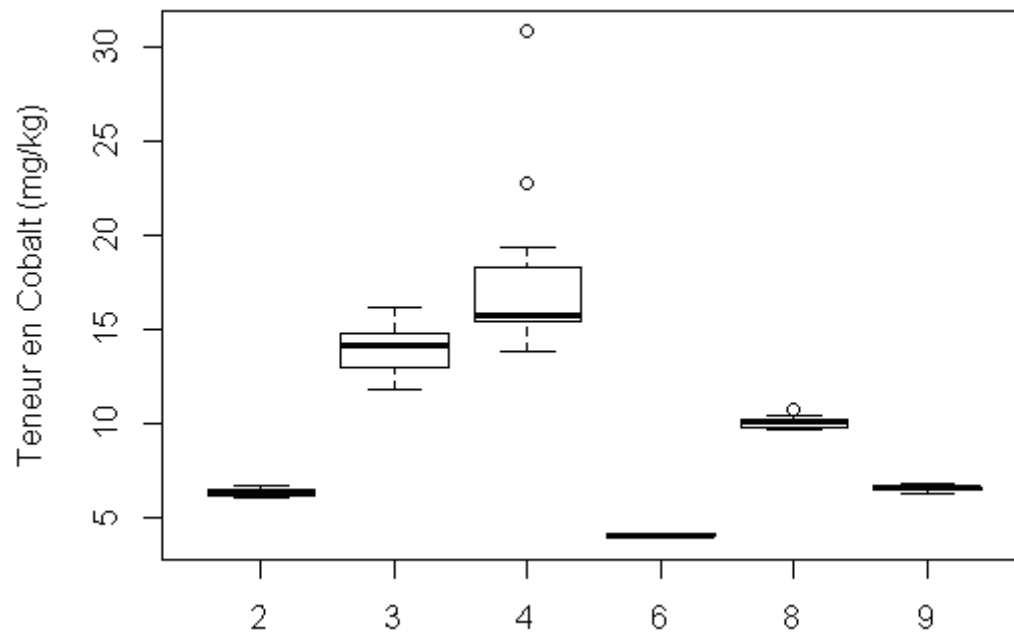
(c) Carbone



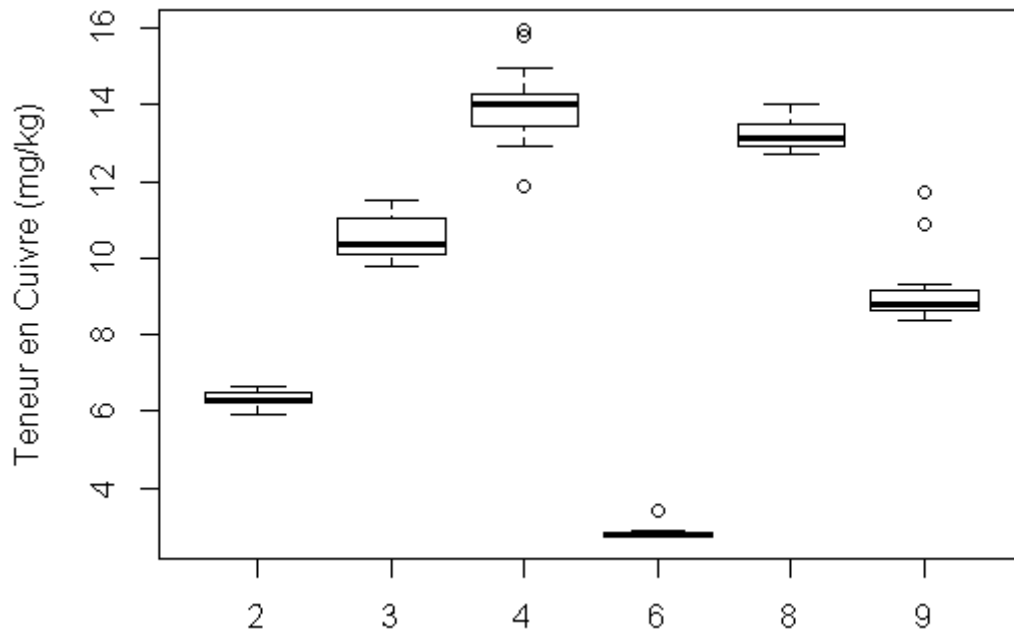
(d) Chrome



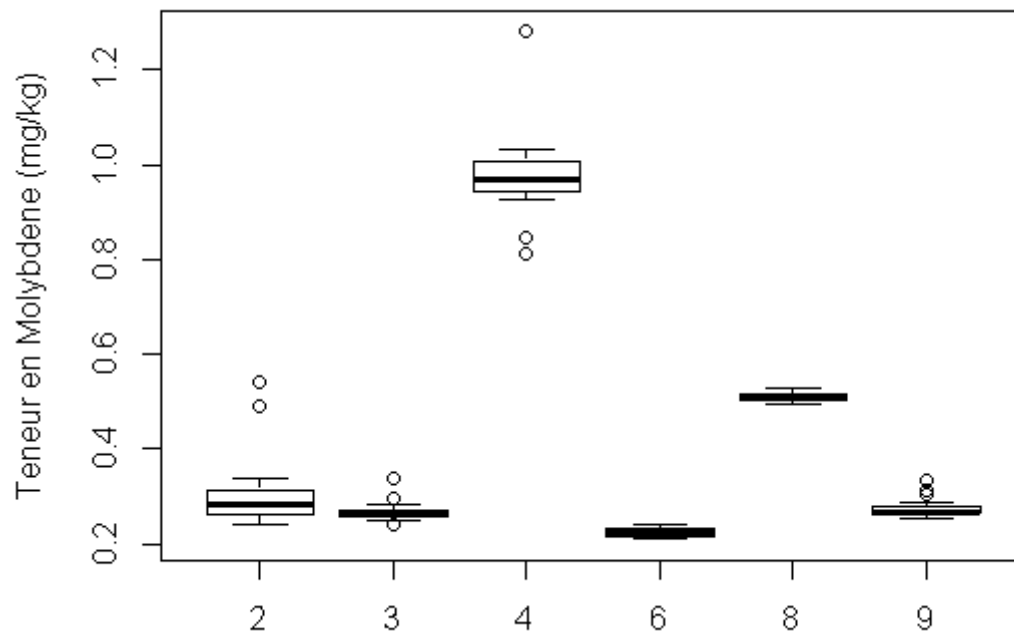
(e) Cobalt



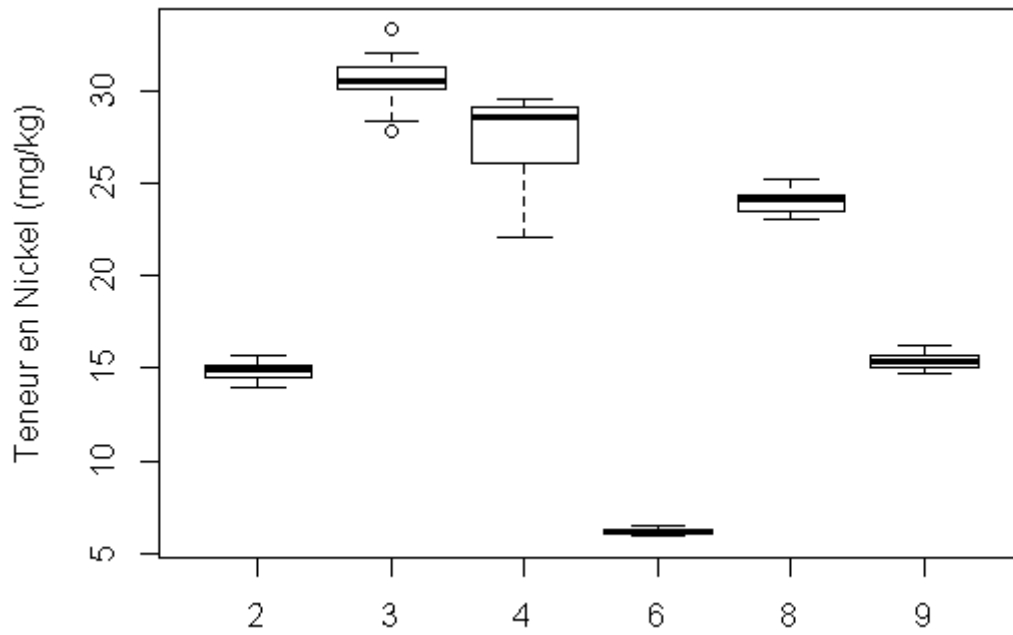
(f) Cuivre



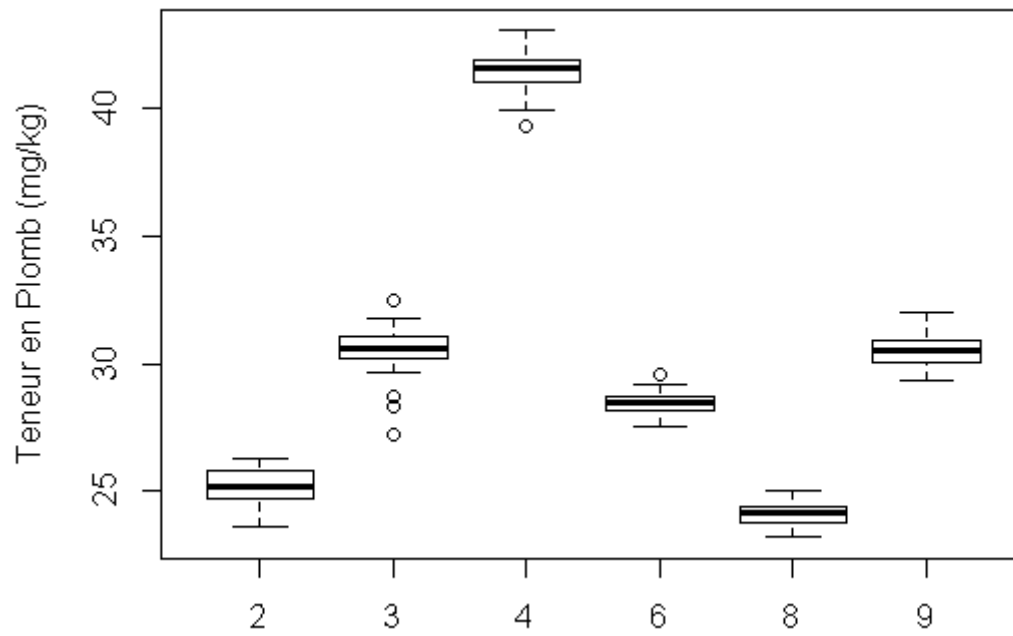
(g) Molybdène



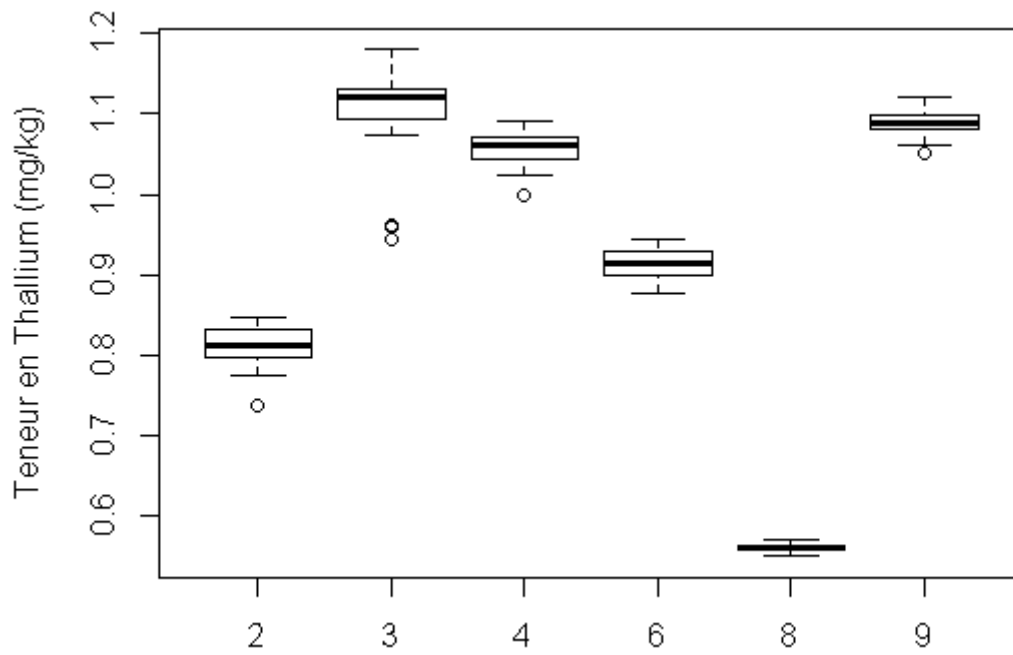
(h) Nickel



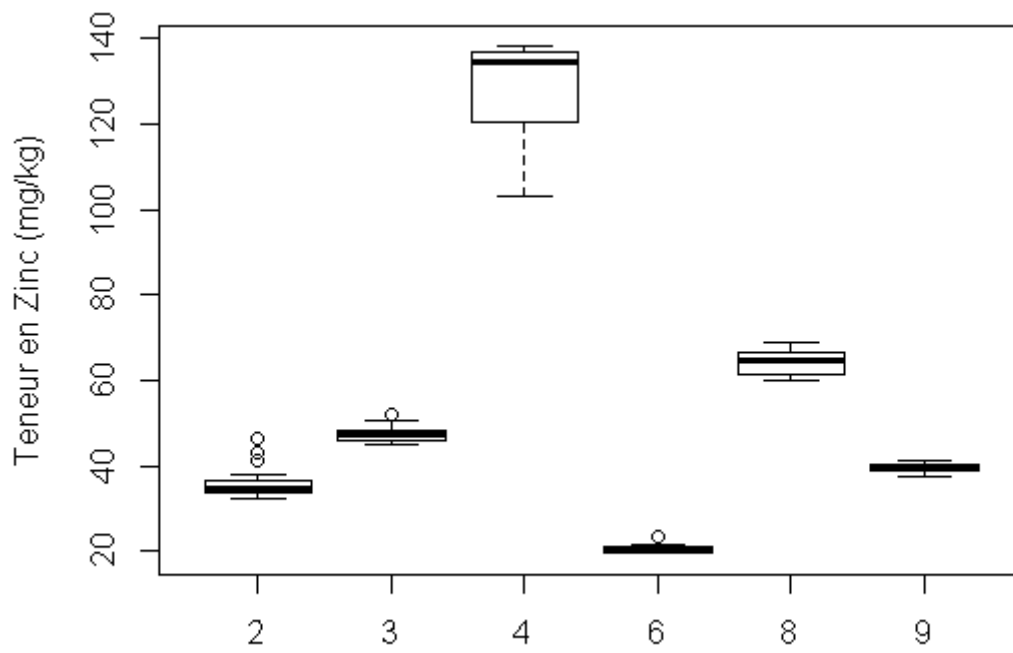
(i) Plomb



(j) *Thallium*



(k) *Zinc*



Annexe 6 : Tableau de comparaisons des modèles mixtes hiérarchique obtenues par test de ratio de vraisemblance et comparaison des AIC, pour chaque élément et pour chaque couche de sol

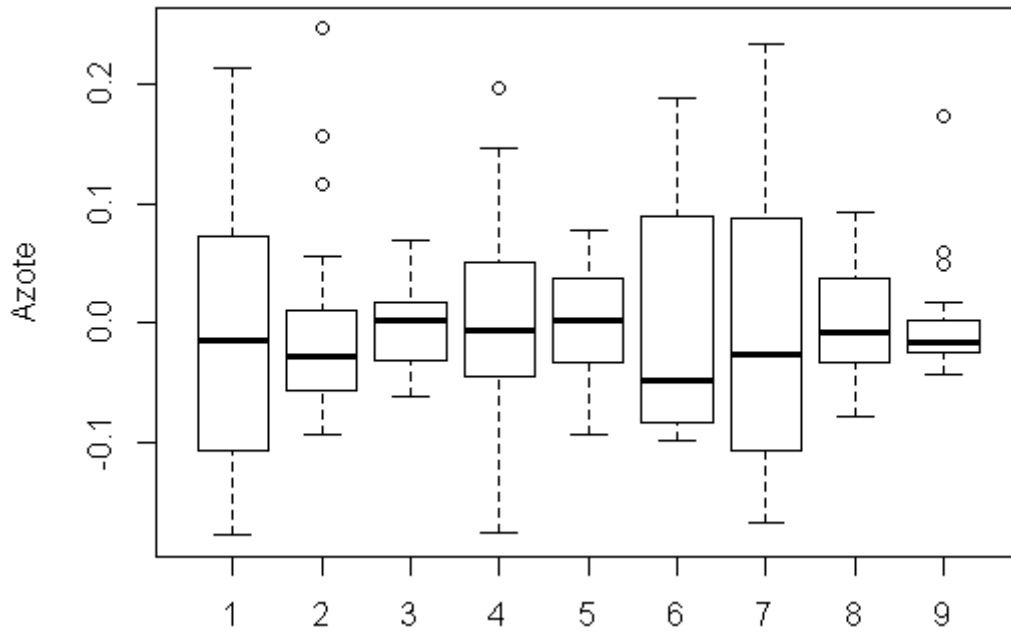
Modèle 1 : uniquement des effets aléatoires hiérarchisés; Modèle 2 : occupation des sols en effet fixe ; Modèle 3 : site en effet fixe

Elément	Horizon	Modèle	df	AIC	logLik	Retenu
Azote	1	1	6	-296.27	154.14	1
		2	9	-293.93	155.97	
		3	14	-362.27	195.14	
	2	1	6	-379.48	195.74	1
		2	8	-379.52	197.76	
		3	11	-414.17	218.08	
Cadmium	1	1	6	-442.03	227.01	1
		2	9	-443.65	230.83	
		3	14	-496.23	262.11	
	2	1	6	-587.63	299.81	3
		2	8	-582.38	299.19	
		3	11	-609.22	315.61	
Chrome	1	1	6	830.5	-409.25	3
		2	9	829.88	-405.94	
		3	14	784.82	-378.41	
	2	1	6	641.99	-314.99	1
		2	8	640.87	-312.43	
		3	11	621.19	-299.59	
Cobalt	1	1	6	291.68	-139.84	2
		2	9	277.49	-129.75	
		3	14	239.83	-105.91	
	2	1	6	364.74	-176.37	1
		2	8	365.41	-174.7	
		3	11	352.34	-165.17	
Cuivre	1	1	6	388.68	-188.34	2
		2	9	381.51	-181.76	
		3	14	341.53	-156.76	
	2	1	6	192.77	-90.39	2
		2	8	187.81	-85.91	
		3	11	167.02	-72.51	
Molybdène	1	1	6	-385.84	198.92	3
		2	9	-387.37	202.69	
		3	14	-438.47	233.24	
	2	1	6	-266.58	139.29	1
		2	8	-263.97	139.98	
		3	11	-295.74	158.87	
Nickel	1	1	6	350.91	-169.46	2
		2	9	347.61	-164.8	
		3	14	283.06	-127.53	
	2	1	6	339.77	-163.89	2
		2	8	335.69	-159.84	
		3	11	310.84	-144.42	

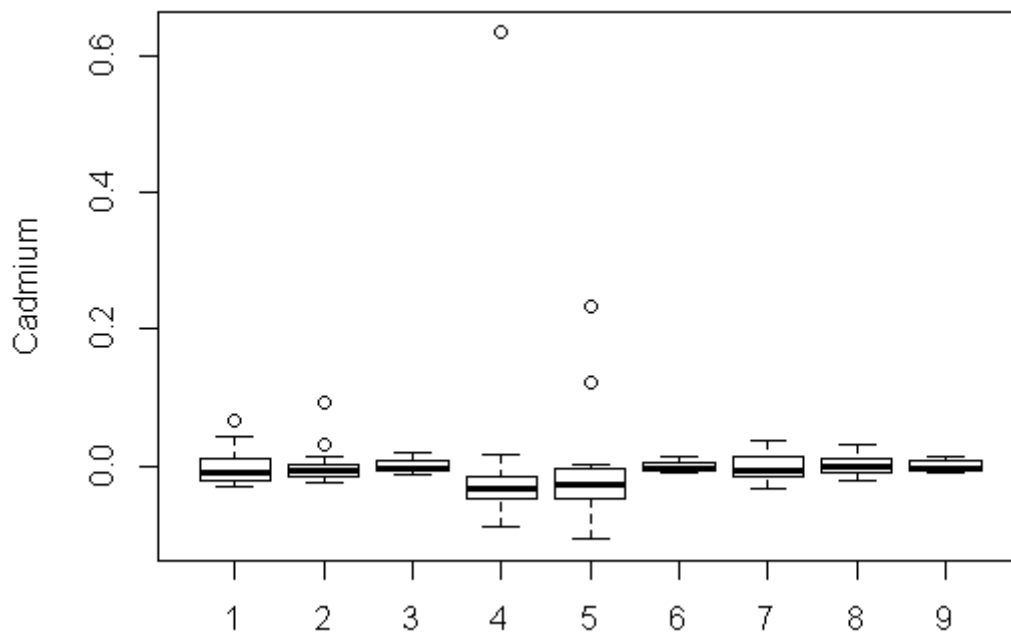
Plomb	1	1	6	555.09	-271.55	3
		2	9	555.43	-268.72	
		3	14	523.97	-247.98	
	2	1	6	284.4	-136.2	1
		2	8	283.26	-133.63	
		3	11	258.03	-118.02	
Thallium	1	1	6	-452.98	232.49	3
		2	9	-451.55	234.78	
		3	14	-488.34	258.17	
	2	1	6	-404.83	208.41	1
		2	8	-403.57	209.78	
		3	11	-432.74	227.37	
Zinc	1	1	6	905.01	-446.5	2
		2	9	901.34	-441.67	
		3	14	841.71	-406.85	
	2	1	6	618.22	-303.11	3
		2	8	618.79	-301.39	
		3	11	586.41	-282.2	

Annexe 7 : Boxplots centrés des teneurs en élément par site de répétabilité pour la couche de surface

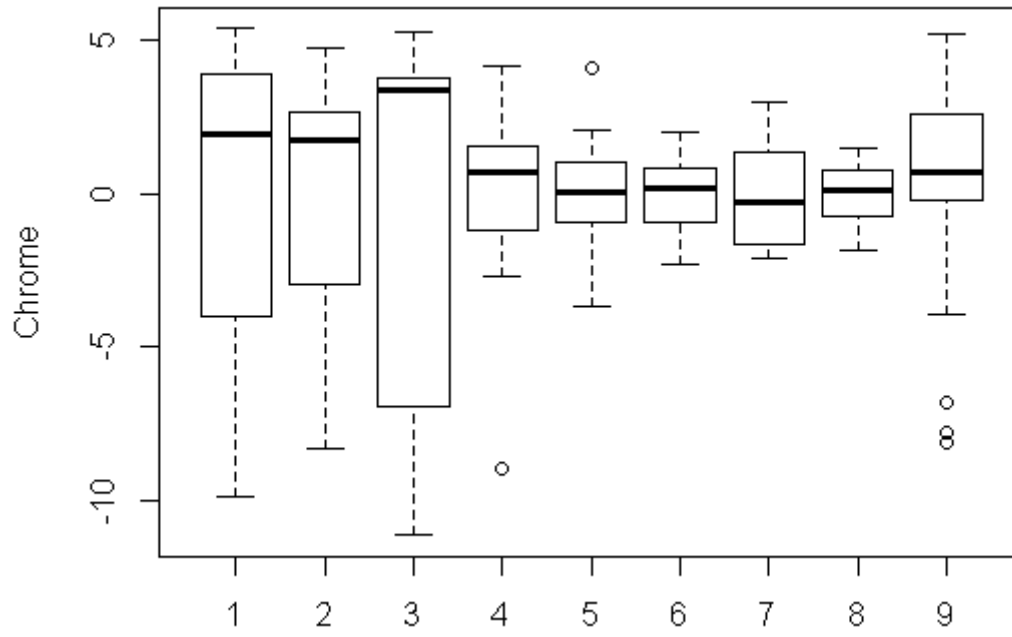
(a) Azote



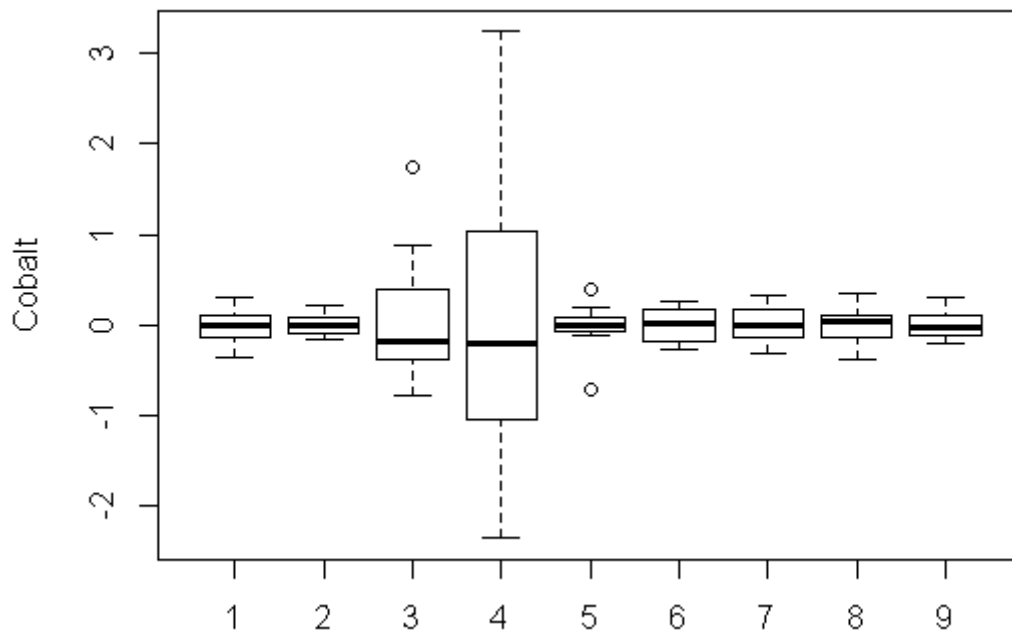
(b) Cadmium



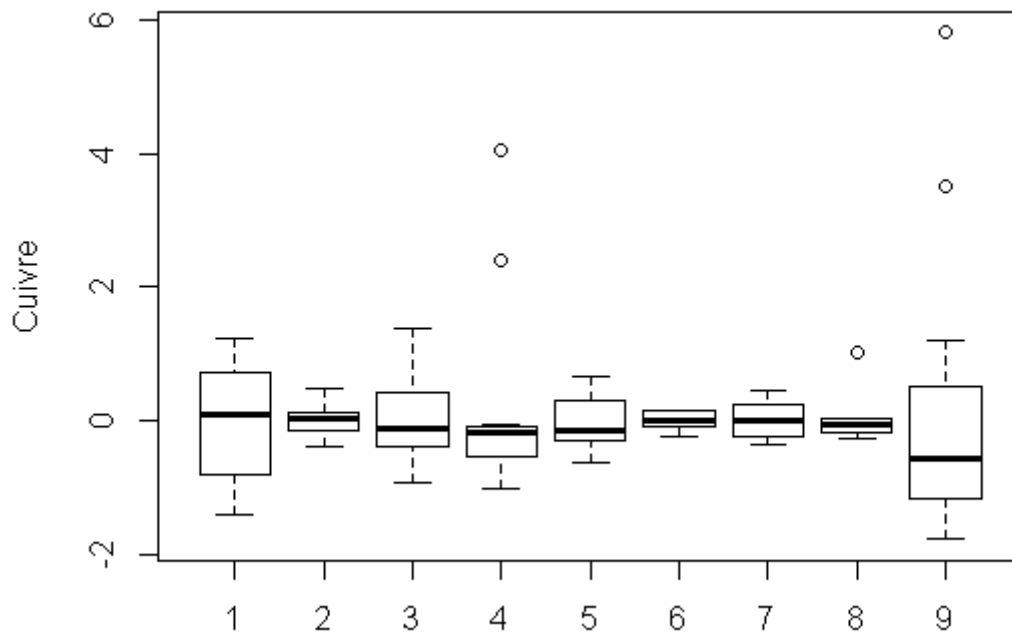
(c) *Chrome*



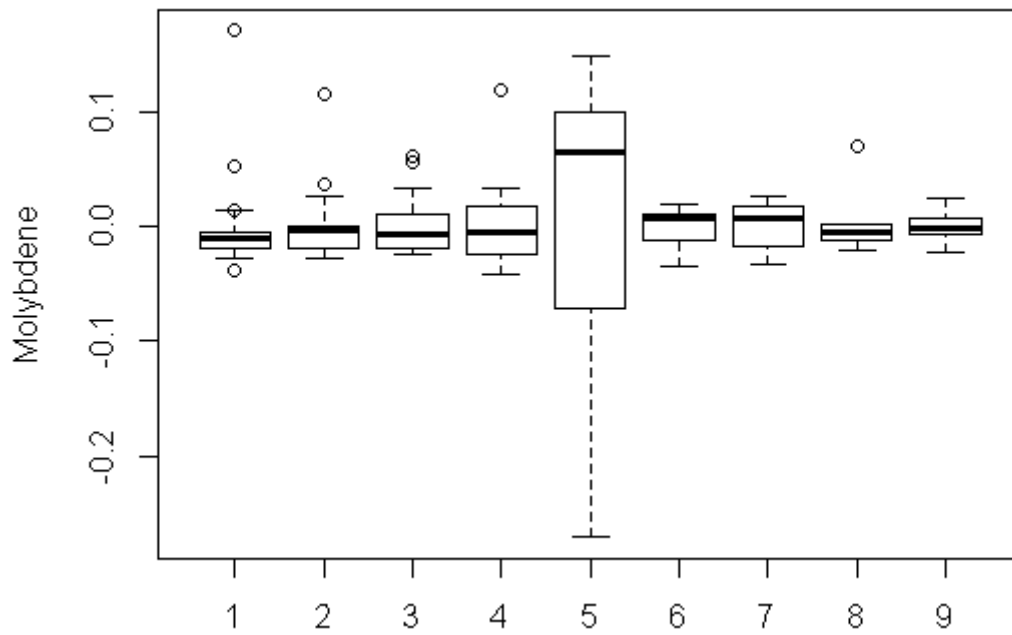
(d) *Cobalt*



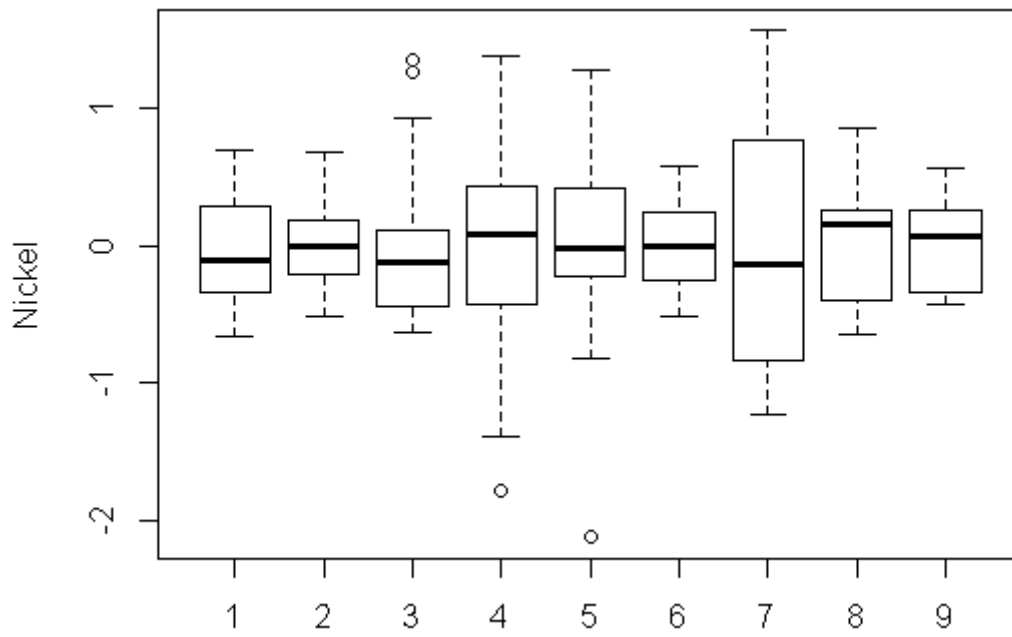
(e) Cuivre



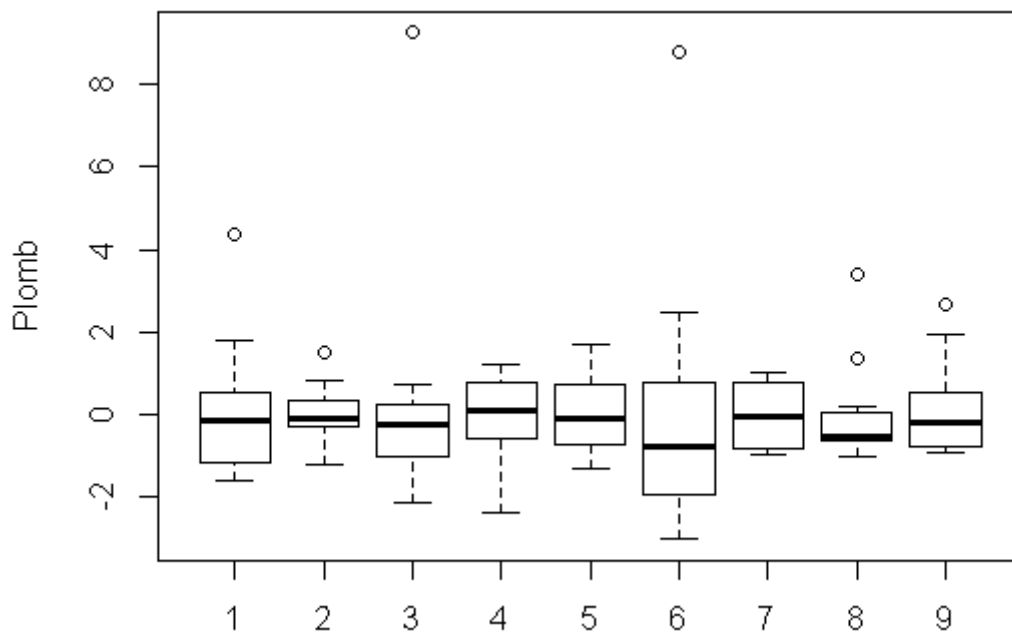
(f) Molybdène



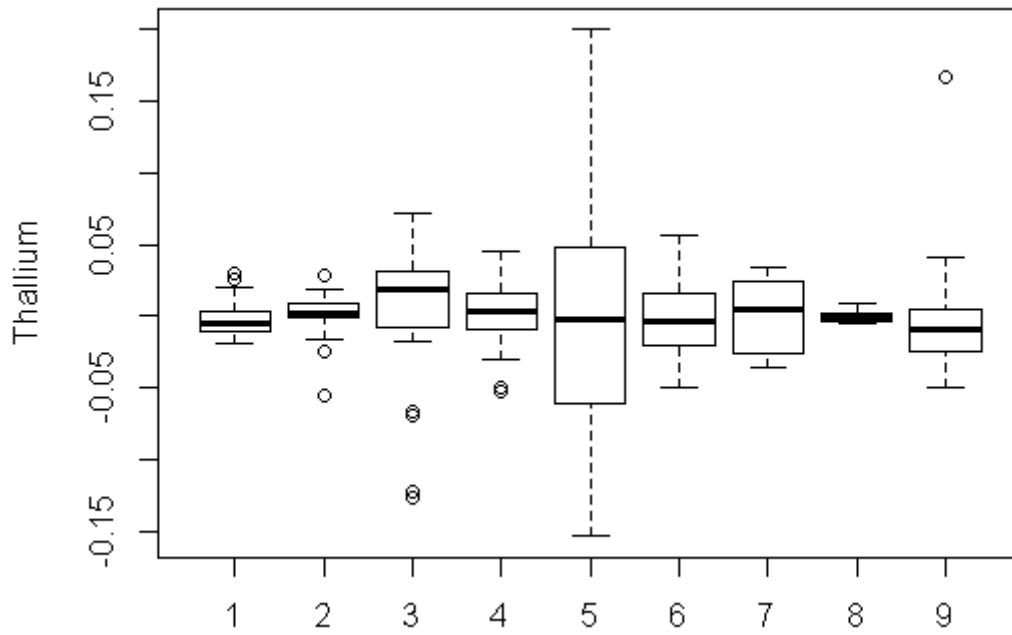
(g) Nickel



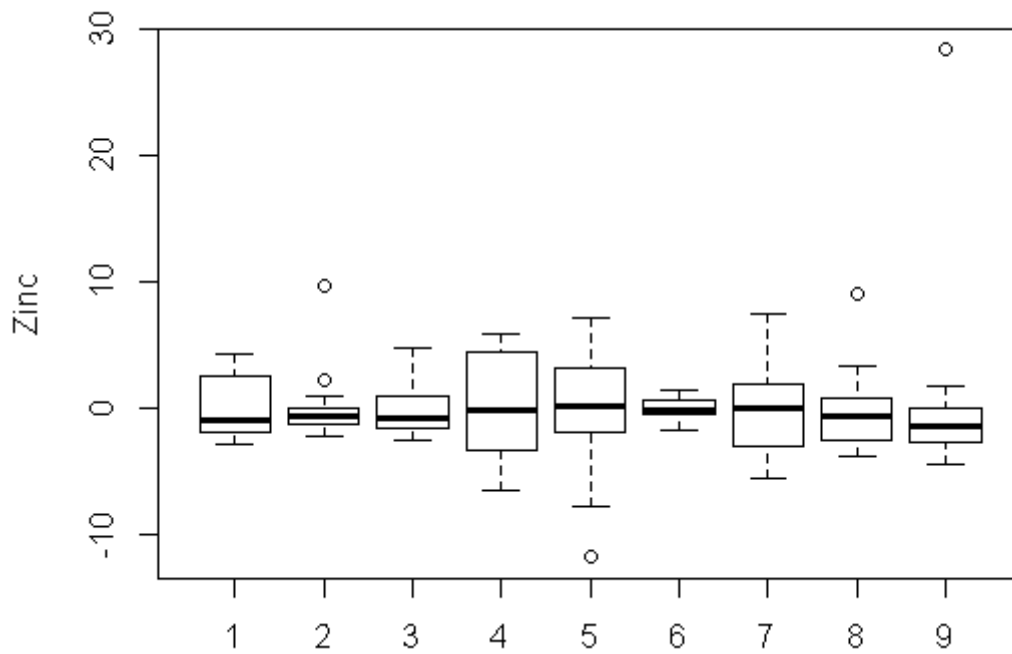
(h) Plomb



(i) *Thallium*

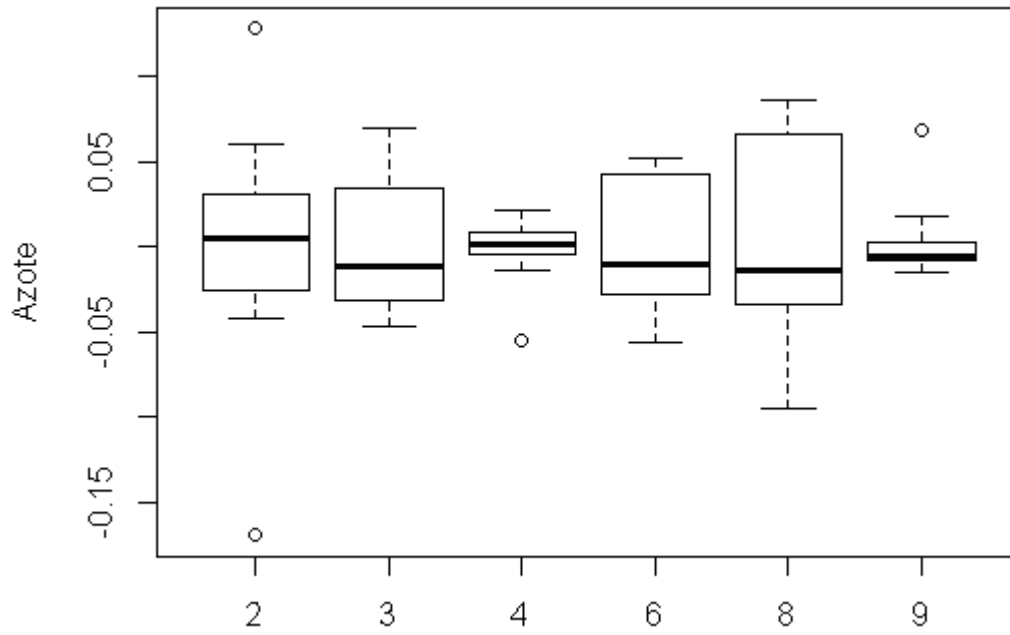


(j) *Zinc*

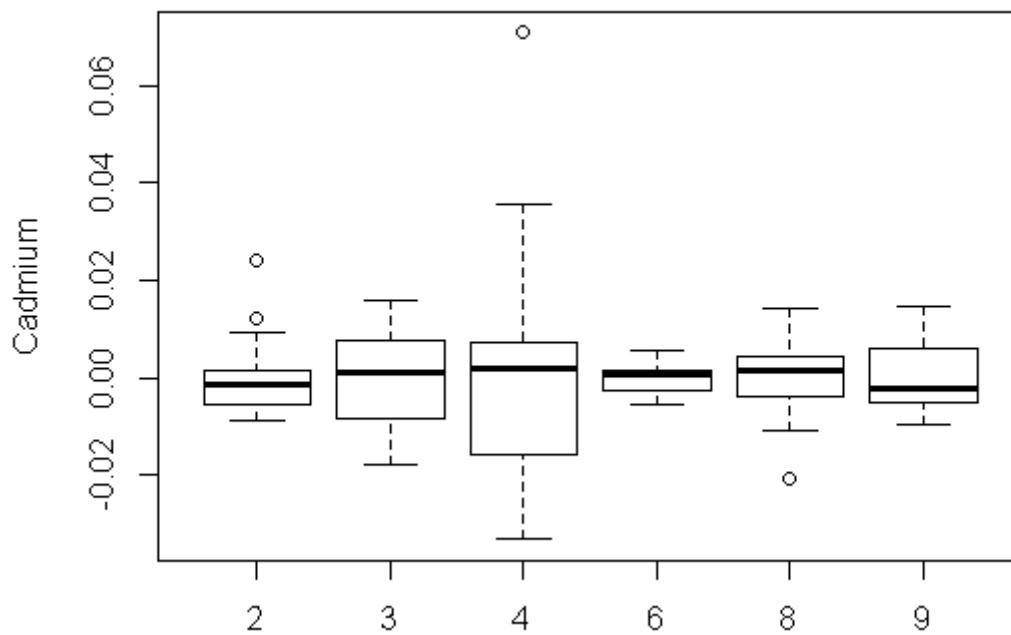


Annexe 8 : Boxplots centrés des teneurs en élément par site de répétabilité pour la couche de sub-surface

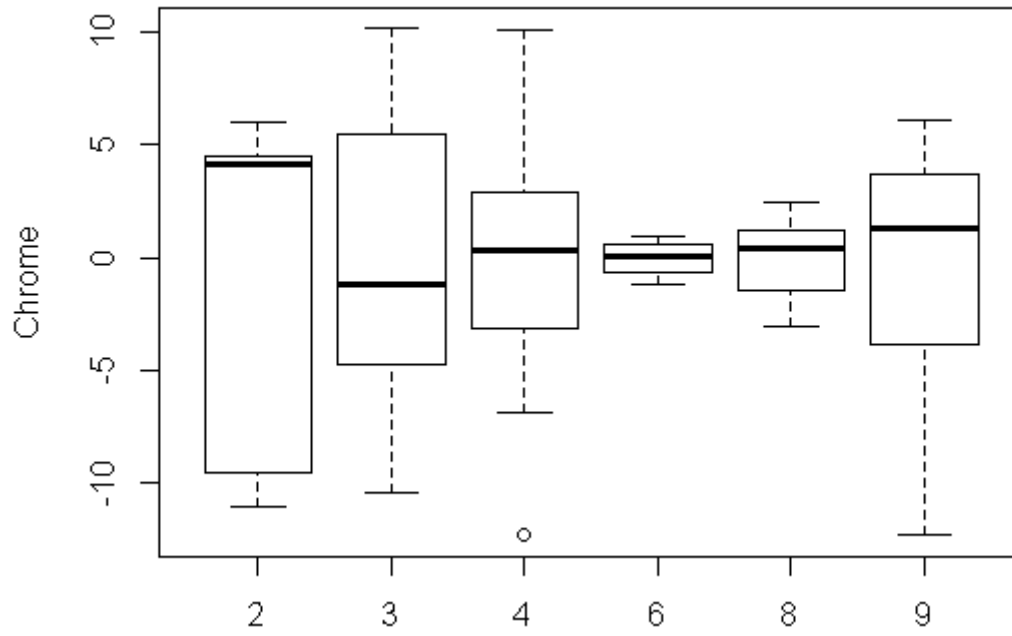
(a) Azote



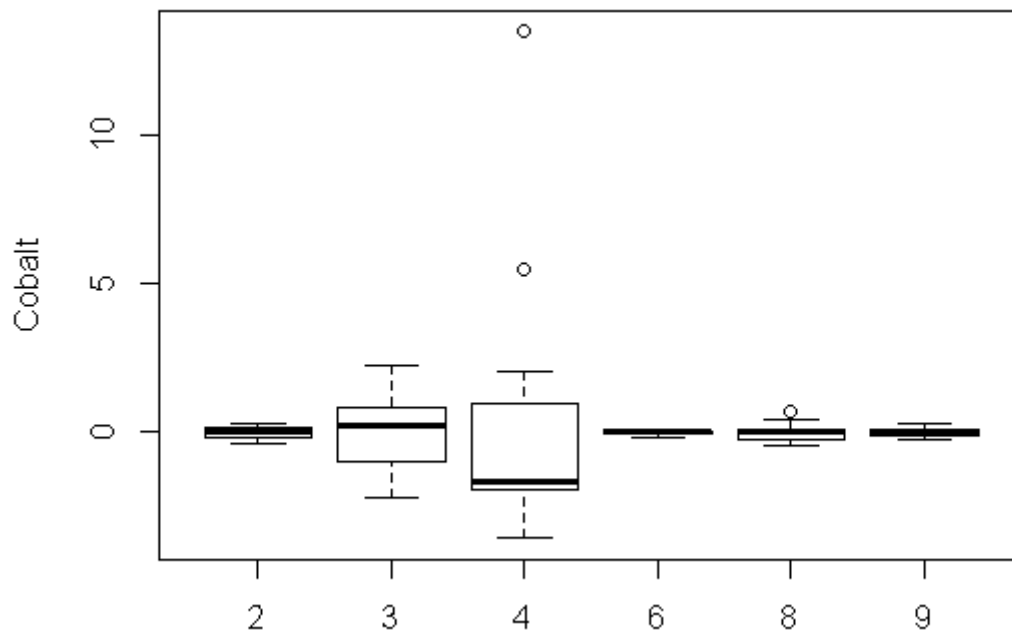
(b) Cadmium



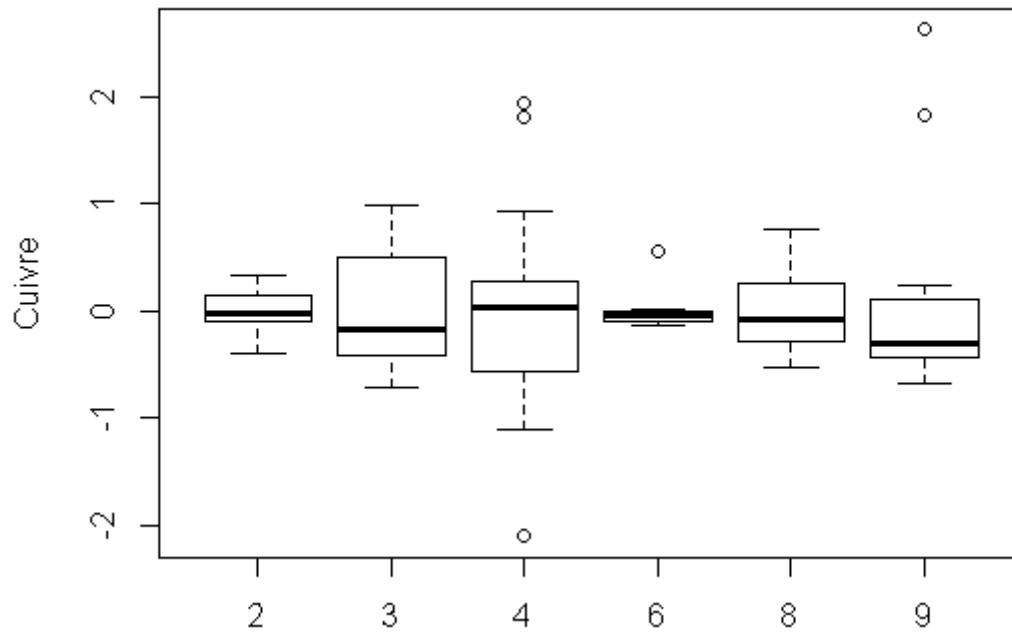
(c) *Chrome*



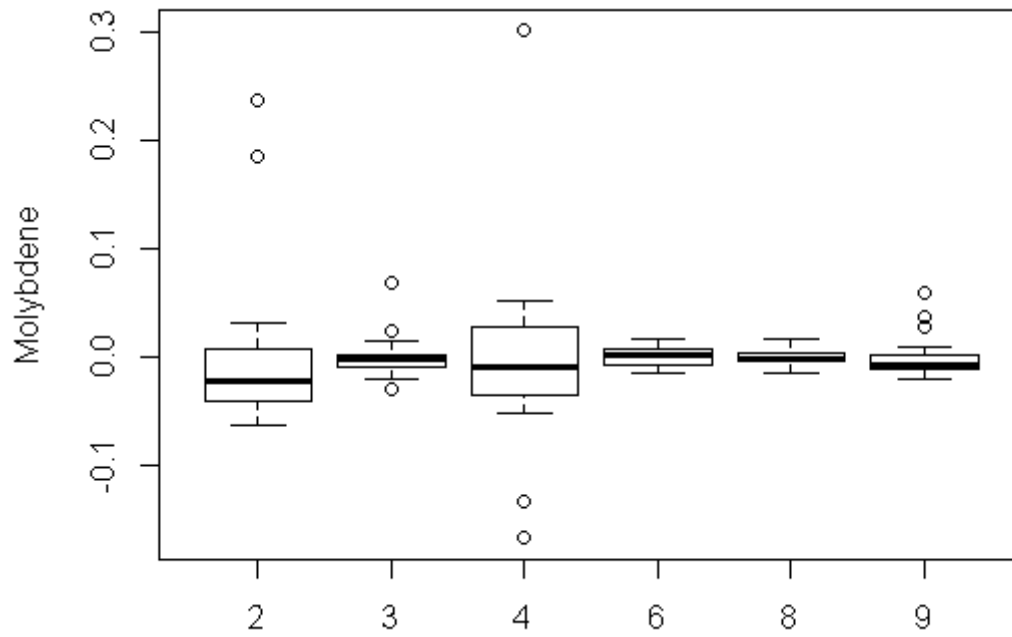
(d) *Cobalt*



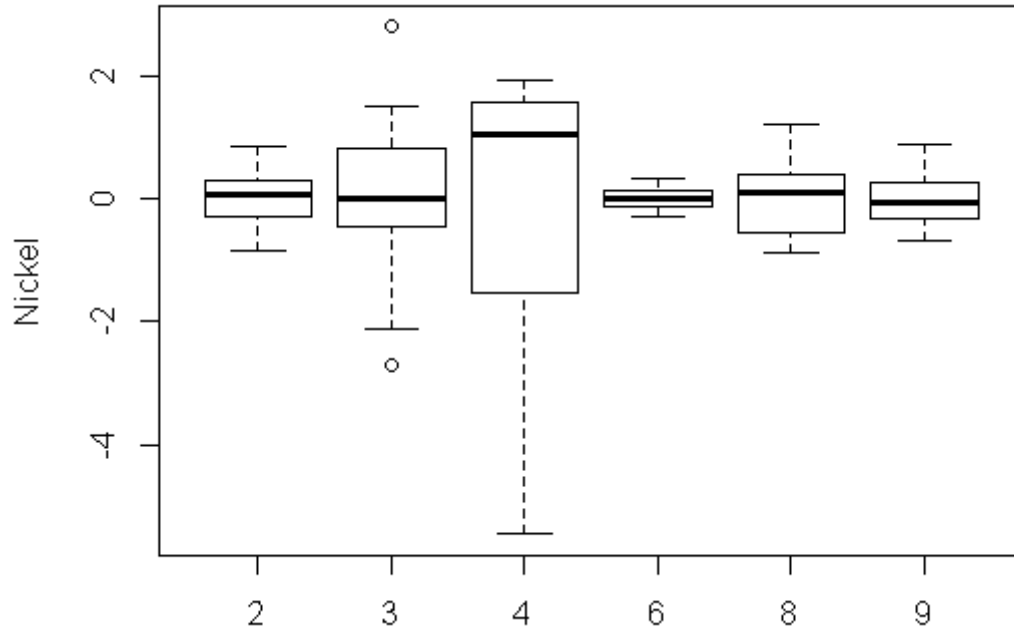
(e) Cuivre



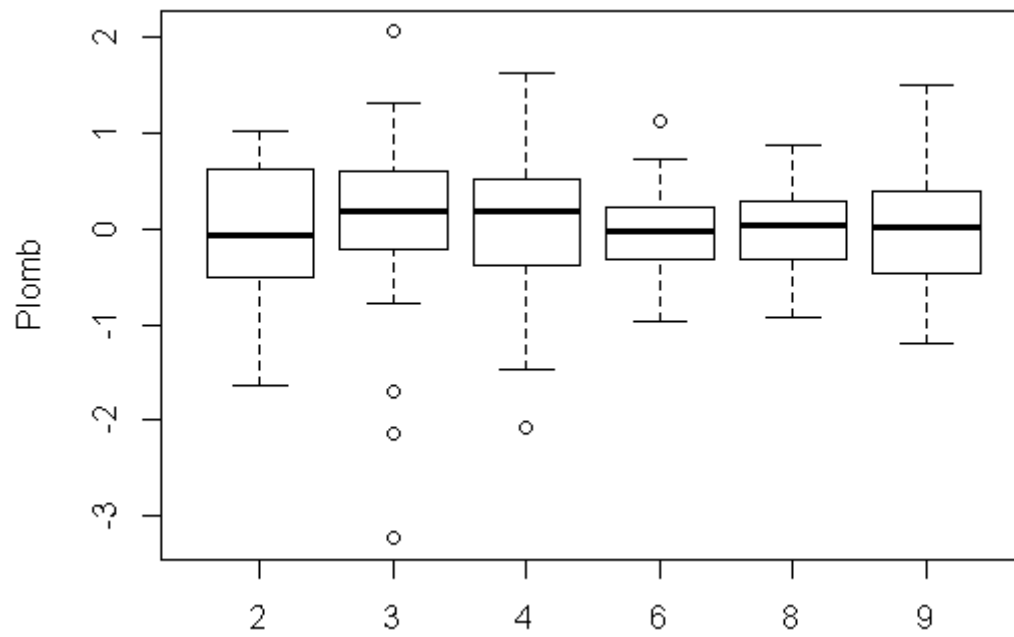
(f) Molybdène



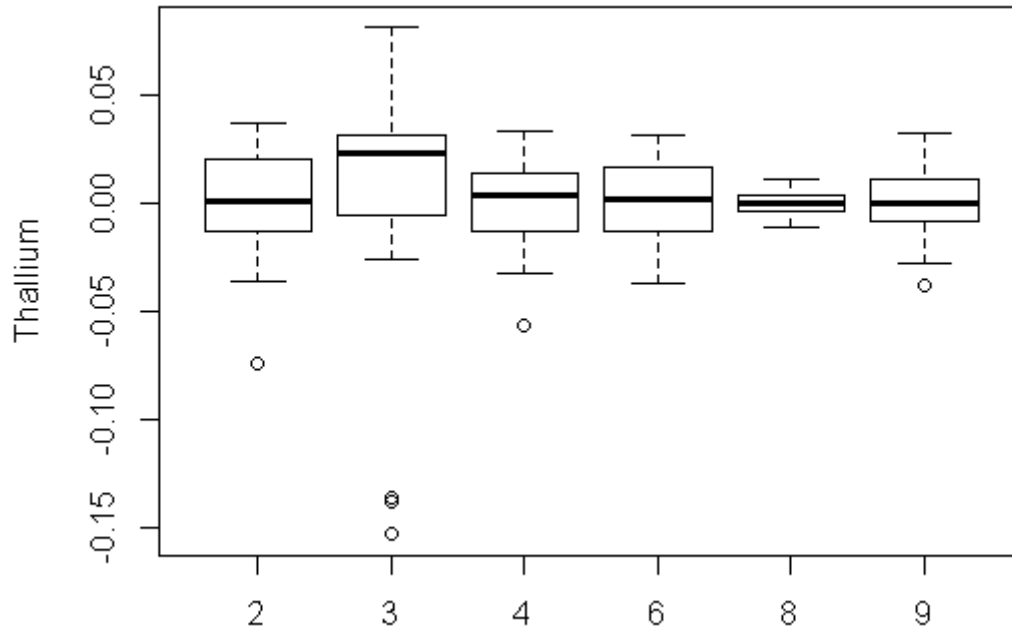
(g) Nickel



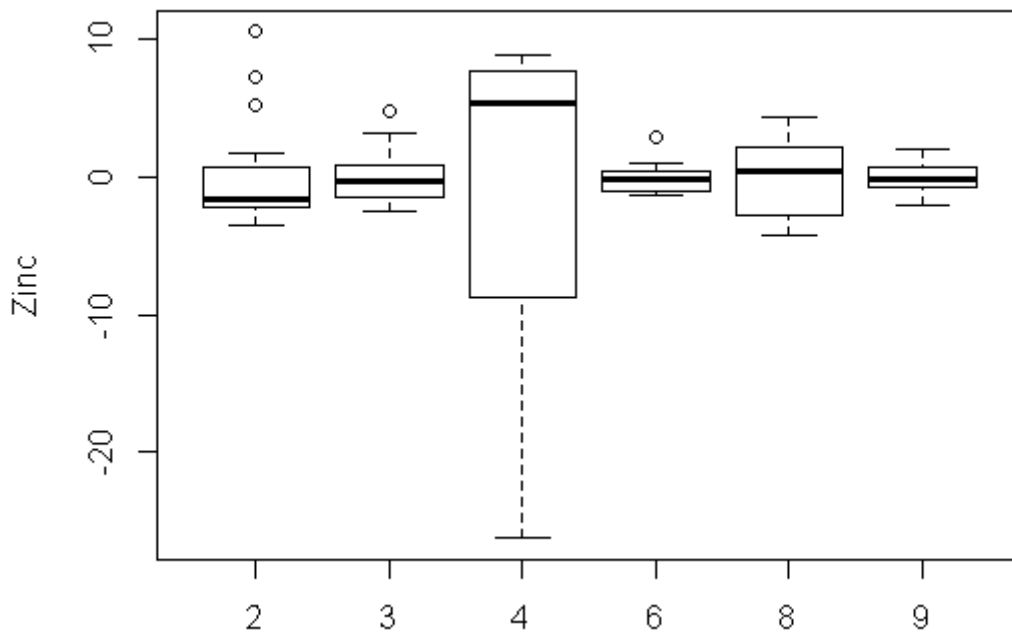
(h) Plomb



(i) *Thallium*



(j) *Zinc*



Annexe 9 : Composants de la variance estimés du modèle qui minimise l'AIC pour la couche de surface

Eléments	σ_{site}	σ_{comp}	σ_{id2}	σ_{id3}	σ_{res}
Azote	1.75	4.3×10^{-2}	4.5×10^{-2}	6.1×10^{-2}	2.6×10^{-2}
Cadmium	8.3×10^{-1}	9.4×10^{-3}	8.9×10^{-4}	6.8×10^{-2}	6.2×10^{-3}
Chrome	2.8×10^2	1.7	2.64	1.38	1.86
Cobalt	2.34	4.7×10^{-1}	1.5×10^{-1}	5.6×10^{-1}	1.1×10^{-1}
Cuivre	7.0	1.0	2.0×10^{-1}	6.2×10^{-1}	2.1×10^{-1}
Molybdène	5.9×10^{-1}	3.3×10^{-2}	1.9×10^{-2}	3.9×10^{-2}	3.1×10^{-2}
Nickel	1.1×10^1	3.3×10^{-1}	2.9×10^{-1}	4.3×10^{-1}	3.2×10^{-1}
Plomb	5.9	8.1×10^{-1}	1.5	9.5×10^{-1}	3.6×10^{-1}
Thallium	2.2×10^{-1}	4.5×10^{-2}	1.0×10^{-2}	4.3×10^{-4}	3.7×10^{-2}
Zinc	8.1×10^1	1.47	3.7	2.1	1.8

Annexe 10 : Composants de la variance estimés du modèle qui minimise l'AIC pour la couche de sub-surface

Eléments	σ_{site}	σ_{comp}	σ_{id2}	σ_{id3}	σ_{res}
Azote	4.6×10^{-1}	2.6×10^{-2}	2.7×10^{-2}	3.0×10^{-2}	1.2×10^{-2}
Cadmium	7.3×10^{-5}	1.5×10^{-5}	9.0×10^{-6}	4.5×10^{-6}	7.7×10^{-6}
Chrome	1.6×10^{-2}	2.3×10^{-3}	4.1×10^{-3}	1.5×10^{-3}	3.1×10^{-3}
Cobalt	6.0×10^{-3}	3.0×10^{-3}	2.5×10^{-4}	2.2×10^{-4}	8.0×10^{-4}
Cuivre	2.1×10^{-3}	6.1×10^{-4}	3.5×10^{-4}	1.9×10^{-4}	3.2×10^{-4}
Molybdène	2.3×10^{-4}	2.0×10^{-5}	2.3×10^{-5}	1.5×10^{-5}	4.9×10^{-5}
Nickel	6.3×10^{-3}	1.7×10^{-4}	1.1×10^{-3}	1.1×10^{-3}	7.0×10^{-4}
Plomb	6.2×10^{-3}	7.9×10^{-4}	3.2×10^{-4}	4.3×10^{-4}	5.2×10^{-4}
Thallium	2.0×10^{-4}	1.2×10^{-5}	3.7×10^{-5}	1.4×10^{-5}	1.5×10^{-5}
Zinc	3.6×10^{-2}	1.6×10^{-3}	4.3×10^{-3}	1.9×10^{-3}	2.2×10^{-3}