



Master 2



Influence des plantes associées aux cultures de riz ou de maïs à Madagascar sur les vers blancs et la biodiversité de la faune du sol

présenté par

Jessica Ambard

Mémoire préparé sous la direction de :

Philippe LETOURMY

Présenté le :

21/06/2010

Organisme d'accueil :

Cirad

Devant le jury :

Yann GUEDON
Christian LAVERGNE
Philippe LETOURMY
Jean-Michel MARIN
Nicolas MOLINARI
Robert SABATIER

CIRAD-DIST
Unité bibliothèque
Lavalette



0000105069

Remerciements

Je souhaite adresser ici tous mes remerciements aux personnes qui m'ont apporté leur aide, et ont ainsi contribué à l'élaboration définitive de ce mémoire.

Tout d'abord, je tiens à exprimer toute ma gratitude envers M. Letourmy, biostatisticien au Cirad et directeur de ce mémoire, pour sa gentillesse et sa bienveillance à mon égard, mais aussi pour sa disponibilité tout au long de ce parcours, malgré son emploi du temps chargé. Son aide m'a été naturellement très précieuse pour rédiger ce rapport, et notamment pour m'initier à la méthodologie que nécessite une telle étude.

Je souhaite également témoigner toute ma reconnaissance à l'équipe de l'ATP Omega 3, pour m'avoir accueillie dans l'un de leurs projets en cours, et permis ainsi de participer à leur propre recherche. En particulier, Mme Rabary, Mme Rafaraso, M. Michellon, M. Naudin, M. Randriamanantsoa, et M. Ratnadass, pour nos nombreux échanges, lesquels m'ont permis de comprendre et d'apprécier le lien étroit entre agronomes et biostatisticiens.

Par ailleurs, je tiens aussi à remercier M. Ducharme, et M. Lavergne, enseignants du Master 2 Biostatistiques, ainsi que Mme Trottier et M. Lancelot, d'avoir pris le temps de répondre à toutes mes questions.

Je voudrais également remercier M. Vercambre, entomologiste, spécialisé dans l'étude des vers blancs, M. Gozé biostatisticien, ainsi que toutes les autres personnes qui travaillent à mes côtés et qui m'ont permis, par leur accueil chaleureux, de me sentir à l'aise et dans des conditions adéquates pour mener à bien mes travaux de recherche.

Enfin, j'aimerais remercier mes amis stagiaires pour leur soutien, et leurs précieux conseils.

Table des matières

Présentation de la structure d'accueil	5
1. Introduction.....	6
2. Matériels et méthodes	8
2.1 Présentation des données	8
2.1.1 Dispositif expérimental	8
2.1.2 Méthodes de mesures	9
2.2 Méthodes de traitements.....	11
2.2.1 Densités de vers blancs.....	11
2.2.2 Diversité de la macrofaune du sol	12
2.2.3 Diversité de la macrofaune des Pitfall trap	13
2.2.4 Production de riz	13
2.2.5 Dégâts sur les cultures de riz.....	13
2.2.6 Effet biocide des associations de plantes sur les larves en laboratoire.....	14
2.3 Modèles linéaires généralisés.....	15
2.3.1 Structure du modèle.....	15
2.3.2 Estimation dans les modèles linéaires généralisés.....	17
2.4 Modèles linéaires mixtes.....	20
2.4.1 Description du modèle.....	20
2.4.2 Moments conditionnels et marginaux.....	20
2.4.3 Autres propriétés.....	20
2.4.4 Vraisemblance	21
2.4.5 Estimation dans les modèles linéaires généralisés mixtes.....	21
2.5 Modèles linéaires généralisés mixtes.....	24
2.5.1 Description du modèle.....	24
2.5.2 Estimations dans les GLMMs.....	24
2.6 Surdispersion	28
2.6.1 Mise en évidence de la surdispersion.....	28
2.6.2 Paramètre de dispersion.....	29
2.6.3 Conséquences de l'ajustement	29
2.7 Logiciels	30
2.7.1 Fonction glmer de R.....	30

2.7.2	Procédure GLIMMIX de SAS.....	30
2.7.3	Comparaison.....	31
3.	Résultats et discussion.....	32
3.1	Protocole suivi pour chaque variable	32
3.2	Application des méthodes présentées aux densités de vers blancs dans le dispositif riz pluvial au mois de mars	32
3.2.1	Analyse exploratoire	32
3.2.2	Présentation du modèle	34
3.2.3	Ajustement	35
3.2.4	Validation.....	38
3.2.5	Analyse des effets fixes.....	38
3.2.6	Estimation des moyennes.....	40
3.2.7	Comparaisons de moyennes.....	40
3.3	Présentation synthétique des résultats et interprétations	41
3.3.1	Densités de vers blancs.....	42
3.3.2	Diversité de la macrofaune du sol et des Pitfall trap.....	43
3.3.3	Production de riz et dégâts observés	44
3.3.4	Effets biocide des associations de plantes sur les larves.....	45
3.4	Limites de ce travail, et suggestions d'améliorations.....	46
4.	Conclusion	47
	Bibliographie.....	49
	Annexes	51

Présentation de la structure d'accueil

- **Le Cirad**

Le Cirad (Centre de coopération Internationale en Recherche Agronomique pour le Développement) est un centre de recherche français qui répond, avec les pays du Sud, aux enjeux internationaux de l'agriculture et du développement. Cet établissement public comporte trois départements scientifiques : Systèmes biologiques (Bios), Performance des systèmes de production et de transformation tropicaux (Persyst), et Environnement et sociétés (ES), qui constituent ensemble plus d'une cinquantaine d'unités de recherche.

A partir de ses directions régionales, le CIRAD mène des activités de coopération avec plus de 90 pays. Il développe par ailleurs des pôles scientifiques à vocation régionale dans l'outre-mer français. En France métropolitaine, il met à la disposition de la communauté scientifique nationale et internationale, un important dispositif de recherche et de formation situé principalement à Montpellier.

En partenariat avec les pays du Sud, le Cirad produit et transmet de nouvelles connaissances, afin d'accompagner leur développement agricole, d'apporter une aide à la décision aux professionnels du secteur de l'agriculture, et de contribuer au débat sur les grands enjeux mondiaux de l'agronomie.

Sa stratégie dépend des mutations économiques, sociales et environnementales engendrées par le phénomène de la mondialisation. Cependant, le CIRAD concentre principalement ses activités autour de six axes scientifiques prioritaires afin de conforter son dispositif de coopération internationale dans le monde :

- l'intensification écologique,
- la biomasse énergie et les sociétés du Sud,
- l'alimentation sûre et diversifiée,
- la santé animale et les maladies,
- les politiques publiques, la pauvreté et les inégalités,
- l'agriculture, l'environnement, la nature et les sociétés.

- **L'UPR « Systèmes de culture annuels »**

Cette Unité Propre de Recherche, du département Persyst, a pour principale ambition de concevoir des systèmes de production performants, durables pour les cultures annuelles, tant du point de vue économique, que du point de vue social et environnemental.

Les atouts majeurs de cette unité sont, tout d'abord, une expertise reconnue, mais également des équipes pluridisciplinaires ainsi que des domaines de compétences très diversifiées : analyse des interactions entre génotype et environnement, bioénergies, entomologie, agronomie, biostatistiques, technologies appliquées aux systèmes de grandes cultures tropicales...

1. Introduction

En milieu tropical, certaines cultures sont particulièrement sensibles aux bio-agresseurs. Toutefois, les méthodes de lutte actuellement mises en place contre ces derniers font le plus souvent appels à des *pesticides* de synthèse, lesquels entraînent des risques non négligeables sur la Santé et l'Environnement.

Devant ce constat, le département Persyst du Cirad (Performances des systèmes de production et de transformation tropicaux), qui s'intéresse aux processus biologiques en interaction avec les conditions environnementales et les interventions techniques, a élaboré l'Action Thématique Programmée intitulée : Omega 3 (Optimisation des Mécanismes Ecologiques de Gestion des bio-Agresseurs pour une Amélioration durable de la productivité des Agrosystèmes).

Cette action consiste en une approche d'intensification écologique pour la gestion durable des bio-agresseurs dans les agrosystèmes tropicaux. Elle a pour objectif l'étude des effets de l'introduction planifiée d'une diversité végétale spécifique (DVS) au cœur des agrosystèmes, comme alternative aux pesticides. Pour réaliser cela, elle s'appuie sur six études qui sont différenciées par le bio-agresseur considéré (insecte, plante, etc.), par les plantes-hôtes (céréales, etc.), ou par la DVS employée. Parmi ces investigations, l'une, plus spécifique, a débuté en 2008, et concerne le dispositif qui vise à contrôler les bio-agresseurs, tels que les vers blancs ou le striga, et qui est actuellement mis en place à Madagascar sur des systèmes de cultures à base de riz ou de maïs.

Et c'est ici que vient s'insérer l'objet de ce stage : l'incidence éventuelle de certaines plantes associées à une culture principale¹ sur ces vers blancs ravageurs, et d'une façon plus générale, sur la biodiversité de la faune du sol. Le but recherché est de comparer les effets des différentes plantes de couvertures sur le contrôle de ce ravageur spécifique afin de pouvoir, par la suite, tenter de modéliser ce mécanisme et de prévoir le résultat d'un choix judicieux d'associations de plantes opérationnelles.

Le problème posé est le suivant :

Les différentes plantes-test, associées à la culture principale, ont-elles une influence sur les densités de vers blancs ?

Pour y répondre, il devient primordial de considérer d'abord le système dans son ensemble, car de nombreux paramètres sont étroitement liés : le comportement des vers blancs, la biodiversité de la faune du sol, le biofonctionnement ainsi que la qualité du terrain, les conditions météorologiques ou saisonnières, les divers effets des associations de plantes (appétants, répulsifs, biocides,...) etc., et bien entendu, le développement harmonieux de la culture principale.

Concernant toutes ces interactions complexes, plusieurs hypothèses peuvent être avancées et formulées.

¹ le riz ou le maïs

La première qui survient ici serait de penser que les systèmes de culture avec semis direct sur couverture végétale (SCV) pourraient permettre, dans certaines conditions, de réduire les dégâts occasionnés par les vers blancs sur les cultures de riz et de maïs. Dans ce cas précis, cet effet pourrait être dû à une répulsivité ou toxicité des plantes de services. Ou alors, au contraire, cela pourrait être dû au fait que les vers blancs sont davantage attirés par les plantes de services et n'ont d'effet néfaste que sur celles-ci.

Par ailleurs, on pourrait supposer que les associations de plantes affectent également l'ensemble de la biodiversité de la faune du sol. On pourrait croire notamment que cette biodiversité, générée par l'introduction de ces nouvelles plantes, induise un changement observable de comportement de ces mêmes vers blancs. En particulier, une plus grande diversité de la macrofaune pourrait être à l'origine d'une diminution de la population de ces derniers.

Enfin il se pourrait, contre toute attente, que les plantes de couvertures, *elles-mêmes*, altèrent le développement de la céréale concernée (riz ou maïs).

Pour tenter de répondre *a minima* à toutes ces questions, plusieurs objectifs ont été fixés au cours de cette présente étude:

- Comparer les effets des plantes de service sur les effectifs de vers blancs, et, d'une façon plus générale, sur la biodiversité de la macrofaune (ceci grâce aux indices de diversité),
- Mettre en relation les résultats obtenus avec les données de rendement de riz, et avec les mesures chiffrées des dommages occasionnés sur cette culture principale,
- Comparer les conclusions de ces deux étapes avec celles qui concernent l'effet direct ou indirect des associations de plantes sur le poids des larves de vers blancs en laboratoire.

Afin d'effectuer ce travail, l'outil statistique essentiel utilisé a été le modèle linéaire généralisé mixte (GLMM).

Dans un premier temps, sont donc exposées les données disponibles, ainsi que les méthodes employées pour les analyser. Puis, sont précisés les concepts théoriques qui se réfèrent à cette étude : modèles linéaires généralisés, modèles linéaires mixtes, GLMMs, ainsi que la modélisation de la surdispersion. Les fonctions des logiciels R et SAS utilisées sont partiellement présentées au cours de cette partie.

Dans un deuxième temps, est détaillée l'application des méthodes présentées précédemment au cas particulier d'une variable. Une présentation synthétique des résultats, couplée à leur interprétation, est ensuite réalisée. Pour terminer, sont évoquées les limites apparues lors de cette recherche parcellaire, et sont alors envisagées plusieurs améliorations possibles.

2. Matériels et méthodes

Dans cette partie, nous présentons dans un premier temps la façon dont les données ont été recueillies. Nous exposons également les variables auxquelles nous nous sommes particulièrement intéressés ainsi que les méthodes choisies pour les traiter. Dans un deuxième temps, sont détaillés les concepts théoriques auxquels ces méthodes se réfèrent, et les outils logiciels nécessaires à cette étude.

2.1 Présentation des données

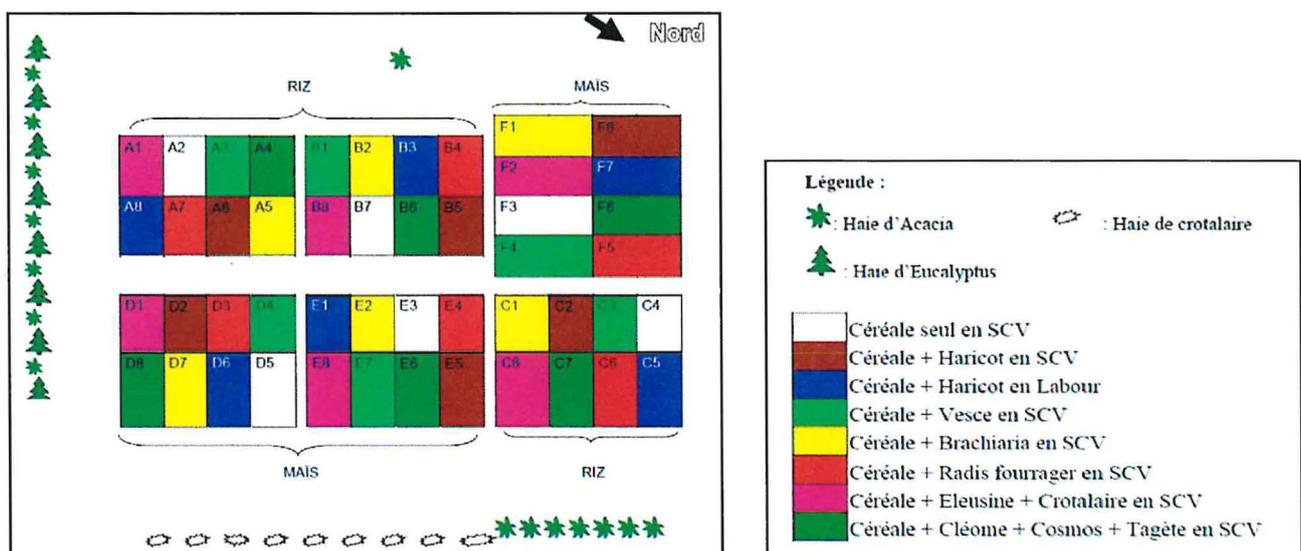
2.1.1 Dispositif expérimental

L'expérience étudiée comportait trois dispositifs :

- dispositif riz pluvial
- dispositif maïs
- dispositif plantes seules

Les dispositifs riz pluvial et maïs étaient situés sur une même grande parcelle de Tafa sur Tanety à sommet plat, et contenaient chacun trois répétitions (blocs) de 8 traitements répartis de façon aléatoire dans chaque bloc.

Parmi l'ensemble des parcelles de l'essai, il y avait notamment des parcelles « témoins » avec les céréales seules, afin de comparer les différences entre céréale seule et céréale avec plantes associées. Mais aussi, des parcelles où la céréale était associée au haricot en SCV pour comparer l'effet des autres plantes de services utilisées au haricot qui a priori n'a pas d'effets sur les ravageurs. Par ailleurs, un traitement céréale associée au haricot en labour a également été mis en place afin de comparer cette fois-ci l'effet du semi direct :



R1: Riz + vesce
R2: Riz + Eleusine + Crotalaire
R3: Riz + Cléome + Tagète + Cosmos
R4: Riz + Brachiaria
R5: Riz + Radis fourrager
RSDH: Riz + Haricot en semis direct
RLH: Riz + Haricot sur Labour
RS: Riz seul
(Rakotoniadranaivo 2009).

M1: Maïs + vesce
M2: Maïs + Eleusine + Crotalaire
M3: Maïs + Cléome + Tagète + Cosmos
M4: Maïs + Brachiaria
M5: Maïs + Radis fourrager
MSDH: Maïs + Haricot en semis direct
MLH: Maïs + Haricot sur Labour
MS : Maïs seul

A, B et C étaient les blocs de riz en 2008-09 (la précédente culture était de l'avoine en 2007-08).
D, E et F étaient les blocs de maïs en 2008-09 (la précédente culture était du riz en 2007-08).

Dans ces deux dispositifs, chaque parcelle élémentaire mesurait 14m de long et 9m de large, espacée de 0.5m entre elles et entre blocs.

Le dispositif plantes seules était situé sur une autre parcelle de TAFE sur flanc de tanety, et composé de 6 traitements en SCV randomisés, avec 3 répétitions (blocs 0, 1 et 2) :

- haricot
- vesce
- brachiaria
- radis fourrager
- crotalaire + eleusine
- cléome + tagète + cosmos

Chaque parcelle élémentaire mesurait 10m de long et 7m de large, et l'espace entre chaque parcelle et chaque bloc était de 0,3m.

Figure 1 (source : Ratnadass et al. 2010)



Figure 2 (source : Ratnadass et al. 2010)



Figure 1 : Dégâts de vers blancs sur une parcelle de riz pluvial, aux semences non traitée aux insecticides.

Figure 2 : Ver blanc en train de consommer des racines de riz pluvial.

2.1.2 Méthodes de mesures

Pour réaliser cette étude, nous avons analysé les données provenant de plusieurs prélèvements différents. Dans un premier temps, un prélèvement de vers blancs uniquement a été réalisé avant la mise en place de

cultures sur les dispositifs riz et maïs, du 11 au 17 novembre 2008, dans le but d'évaluer la population de vers blancs initiales. Dans un deuxième temps, deux prélèvements de toutes les macrofaunes telluriques confondues ont été effectués au cours des mois de mars (saison des pluies) et d'août 2009 (saison sèche).

Les échantillons de sol ont été prélevés selon la méthode standard TSBF (Tropical Soil Biology and Fertility Programme) sur chaque parcelle. Cette méthode est basée sur la technique « bêchage-triage », tri direct des échantillons de sol après bêchage. Chaque échantillon a été prélevé sur un monolithe de dimension 25cm x 25cm x 30cm. Les tris ont été réalisés suivant les quatre horizons du sol suivants: litière, et profondeurs de 0 à 10 cm, 10 à 20 cm et 20 à 30 cm. (Rakotoniadranaivo 2009).

En outre, au cours des mois de mars et d'août 2009, les insectes présents dans les Pitfall trap ont été comptés. Il s'agissait de deux trous à pièges pour chaque parcelle, placés à 1.5m de chaque côté du monolithe. Ils étaient destinés principalement à capturer les fourmis et les arthropodes actifs à la surface du sol. Les pots placés dans le sol avaient un diamètre compris entre 10 et 20cm. Une solution saline et quelques gouttes de détergent étaient versées au fond des pots pour piéger les faunes. Les pièges ont été prélevés après une durée de 48h.

Par ailleurs, pour évaluer la répartition spatiale et temporelle des attaques, les touffes de riz totales et celles attaquées ont été dénombrées selon une notation de 1 à 5 (de « pas attaquées » jusqu'à « totalement attaquées »).

Le comptage a été effectué dans un carré de 2x2m, situé au centre de chaque parcelle élémentaire et a été réalisé tous les 15 jours de la levée jusqu'au stade épiaison de la culture.

De plus, dans le but de procéder à l'évaluation du rendement des cultures de riz, la récolte a été effectuée par parcelle élémentaire dans le carré de 2 m x 2 m où le comptage des attaques a été réalisé.

Enfin, parallèlement et à la même période que ces essais effectués « sur le terrain », l'étude des effets biocides des plantes de couvertures a été réalisée en laboratoire. L'expérience consistait à mesurer le poids de larves de vers blancs élevées dans les six milieux suivants :

- Sol pauvre d'Ibity + paille de riz à 0,5g pour 12g de sol (traitement témoin)
- Sol pauvre d'Ibity + paille de riz à 0,25g + résidus de Radis fourrager à 0,25g pour 12g de sol,
- Sol pauvre d'Ibity + paille de riz à 0,25g + résidus de Vesce à 0,25g pour 12g de sol,
- Sol pauvre d'Ibity + paille de riz à 0,25g + résidus de Desmodium à 0,25g pour 12g de sol,
- Sol pauvre d'Ibity + paille de riz à 0,25g + résidus d'Eleusine à 0,25g pour 12g de sol,
- Sol pauvre d'Ibity + paille de riz à 0,25g + résidus de Cajanus à 0,25g pour 12g de sol.

Ces six traitements étaient répartis en randomisation totale, avec quatre répétitions par traitement, et plusieurs essais ont été effectués sur les trois sortes de larves suivantes : *Heteronychus Arator* (2 essais), *Apicencya Waterlotii* (2 essais), *Heteroconus paradoxus* (3 essais).

Le milieu était changé toutes les semaines et le poids de chaque larve était mesuré au début de l'essai, puis à chaque changement. Les pesées ont été relevées pendant 7 à 8 semaines, ou cessé à la mort des larves.

Ces divers éléments ont permis de déduire les méthodes statistiques adaptées aux données recueillies et décrites dans la section qui suit.

2.2 Méthodes de traitements

2.2.1 Densités de vers blancs

Les densités de vers blancs ont été exprimées en nombre d'individus. Il s'agit donc de données de comptage. Le modèle le plus en adéquation avec ce type de données est le modèle linéaire généralisé, dans ce cas une régression poissonnienne. Toutefois, en étudiant les données de densités de plus près, on s'est aperçu qu'elles étaient surdispersées (cf 2.6), c'est pourquoi nous avons opté plutôt pour une distribution de **quasi-poisson**.

Par ailleurs, d'après le protocole, on se trouvait dans le cas d'un dispositif à deux facteurs : traitement et horizon. Les traitements ont été randomisés dans les blocs (ou grande parcelle) et chaque bloc a été subdivisé en quatre « petites parcelles » (quatre étant le nombre de modalités du facteur horizon). Pour pouvoir étudier un tel dispositif, s'apparentant à un « **split plot** », nous avons considéré le modèle linéaire généralisé mixte suivant :

$\forall i, j, k$ avec $i = 1, \dots, 8$; $j = 1, 2$ et $k = 1, \dots, 3$:

$$\log(\mu_{ijk}) = \nu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + B_{ik}$$

Où:

- $y_{ijk} = \mu_{ijk} + \varepsilon_{ijk}$ est la densité de vers blancs de l'horizon j dans la parcelle i, k ,
- ν est le paramètre de la moyenne générale, estimé par $y_{...}$,
- α_i est l'effet fixe du traitement i , estimé par $y_{i..} - y_{...}$,
- β_j est l'effet fixe de l'horizon j , estimé par $y_{.j.} - y_{...}$,
- γ_k est l'effet fixe du bloc k , estimé par $y_{..k} - y_{...}$,
- $(\alpha\beta)_{ij}$ est l'effet fixe de l'interaction du traitement i avec l'horizon j estimé par $y_{ij.} - y_{i..} - y_{.j.} + y_{...}$,
- B_{ik} est l'effet aléatoire de l'interaction du traitement i avec le bloc k ,
- ε_{ijk} est l'erreur résiduelle de l'horizon j dans la parcelle i, k , estimée par $y_{...}$.

Etant donné que la distribution choisie est quasi-poissonnienne, on a les propriétés suivantes :

$$\mathbb{E}(y_{ijk} | B_{ik}) = \mu_{ijk}$$

$$\text{Var}(y_{ijk} | B_{ik}) = V(\mu_{ijk}) a(\phi) = \mu_{ijk} a(\phi)$$

où $a(\phi)$ est le paramètre de dispersion.

Dans toute la suite, nous utiliserons T pour désigné le facteur traitement, H pour le facteur horizon, et T : H pour désigné l'interaction entre ces deux facteurs.

On en déduit le tableau suivant d'analyse de variance à 2 étages, avec d'une part l'analyse du facteur T (grandes parcelles), et d'autre part l'analyse de l'interaction T : H et du facteur H (« petites parcelles ») présent en annexe (cf *Tableau 1*).

Ce type de dispositif comporte un inconvénient statistique : la précision de la mesure des effets du facteur T est sacrifiée au profit de celle du facteur H. En effet, les mesures de l'effet du facteur H et de l'interaction T : H sont plus précises que celles qui peuvent être obtenues avec un plan en blocs aléatoires complets. Tandis qu'au contraire, la mesure des effets de T est moins précise que celle que l'on obtiendrait avec un plan en blocs aléatoires complets (Jayaraman 1999 p110). Toutefois, nous ne pouvions pas choisir ici un plan en blocs aléatoires complets étant donné l'impossibilité de randomiser les horizons.

2.2.2 Diversité de la macrofaune du sol

Les densités des macrofaunes étaient également exprimées en nombre d'individus. Pour évaluer la biodiversité de cette faune du sol, nous avons utilisé les trois indices de diversité suivants :

Indice de richesse : $\mathcal{H}_r = S - 1$

Cet indice, le plus fréquemment utilisé, est égal au nombre S de catégories (ici, les espèces) diminué de 1, afin qu'une collection comprenant une seule catégorie ait une biodiversité nulle.

Pour faciliter le recours au modèle log-linéaire, nous avons toutefois utilisé ici un indice de richesse exactement égal au nombre d'espèces. Ainsi, un indice de richesse égal à 1 correspond à une diversité nulle.

Indice de Shannon : $\mathcal{H}_s(p) = - \sum_k p_k \ln(p_k)$

Où $p = (p_1, \dots, p_k, \dots)$ est la distribution de fréquences des catégories.

Ceci mesure la perte d'information due à la perte d'une entité.

Indice de Gini-Simpson : $\mathcal{H}_{g.s}(p) = 1 - \sum_k p_k^2$

Cet indice est égal à la probabilité de tirer, avec remise, dans une collection, deux entités appartenant à deux catégories différentes.

(Pavoine 2005).

*Dans la suite, nous avons admis que lorsqu'aucune macrofaune n'est observée (toutes les espèces confondues), la diversité est nulle.

Pour l'étape de modélisation, nous avons utilisé le modèle linéaire généralisé mixte suivant :

$\forall i, j, k$ avec $i = 1, \dots, 8$; $j = 1, 2$ et $k = 1, \dots, 3$:

$$\boxed{g(\mu_{ijk}) = \nu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + B_{ik}}$$

où $\nu, \alpha_i, \beta_j, \gamma_k, (\alpha\beta)_{ij}$, et B_{ik} sont définis comme précédemment.

$y_{ijk} = \mu_{ijk} + \varepsilon_{ijk}$ est cette fois-ci égal à l'indice de diversité de l'horizon j de la parcelle i, k .

L'indice de richesse, a été modélisé par une distribution **poissonnienne** (car les données n'étaient pas surdispersées) avec une fonction de lien logarithmique.

Tandis que les indices de Shannon et de Gini-Simpson ont été modélisés par une distribution **gaussienne** avec pour fonction de lien la fonction identité.

2.2.3 Diversité de la macrofaune des Pitfall trap

Les données étaient cette fois encore exprimées en nombre d'individus. L'analyse de la biodiversité de la macrofaune dans les Pitfall trap a été effectuée au moyen des indices de diversité énoncés précédemment. Étant donné le dispositif expérimental mis en place pour recueillir les données des Pitfall trap, nous avons choisi d'étudier le modèle linéaire généralisé suivant :

$\forall i, k$ avec $i = 1, \dots, 8$ et $k = 1, \dots, 3$:

$$g(\mu_{ik}) = \nu + \alpha_i + \gamma_k$$

où $y_{ik} = \mu_{ik} + \varepsilon_{ik}$ est l'indice de diversité de la parcelle i, k , et ν, α_i , et γ_k sont définis comme précédemment,

Pour l'indice de richesse, on a choisi une distribution **poissonienne** avec une fonction de lien log, et pour les indices de Shannon et de Gini-Simpson, on a utilisé une distribution **gaussienne** avec la fonction de lien identité.

2.2.4 Production de riz

Les données de production, exprimées en kilogramme par parcelle élémentaire, ont été traitées avec le modèle linéaire suivant:

$\forall i, k$ avec $i = 1, \dots, 8$ et $k = 1, \dots, 3$:

$$y_{ik} = \nu + \alpha_i + \gamma_k + \varepsilon_{ik}$$

Où :

- $y_{ik} = \mu_{ik} + \varepsilon_{ik}$ est la production de riz de la parcelle i, k ,
- ν, α_i , et γ_k sont définis comme précédemment, et
- $\varepsilon_{ik} \sim \mathcal{N}(0, \sigma^2 I)$ est l'erreur résiduelle de la parcelle i, k .

2.2.5 Dégâts sur les cultures de riz

Pour examiner les données de poquets de riz attaqués, nous avons choisi d'étudier le taux de poquets attaqués sur les poquets restants. Cette variable étant exprimée en pourcentage, la distribution adéquate est la distribution binomiale. Cependant, comme les données étaient surdispersées, nous avons choisi une distribution quasi-binomiale. Le modèle est donc le suivant:

$\forall i, l, k$ avec $i = 1, \dots, 8$; $l = 1, \dots, 5$ et $k = 1, \dots, 3$:

$$\boxed{\text{logit}(\mu_{ilk}) = \nu + \alpha_i + \delta_l + \gamma_k + (\alpha\delta)_{il} + B_{ik}}$$

Où :

- $y_{ilk} = \mu_{ilk} + \varepsilon_{ilk}$ est le taux de poquets attaqués sur le nombre de poquets restants à la date l dans la parcelle i, k
- ν, α_i, γ_k , et B_{ik} sont définis comme précédemment,
- δ_l est l'effet fixe de la date l (levée, début tallage, mi-tallage, montaison, épiaison)
- $(\alpha\delta)_{il}$ est l'effet fixe de l'interaction du traitement i avec la date l ,
- ε_{ilk} est l'erreur résiduelle à la date l dans la parcelle i, k .

On a les propriétés suivantes :

$$\mathbb{E}(y_{ijk} | B_{ik}) = \mu_{ijk}$$

$$\text{Var}(y_{ijk} | B_{ik}) = V(\mu_{ijk}) a(\phi) = \mu_{ijk} (1 - \mu_{ijk}) a(\phi)$$

où $a(\phi)$ est le paramètre de dispersion.

2.2.6 Effet biocide des associations de plantes sur les larves en laboratoire

Les poids des larves sont mesurés en grammes.

Nous avons choisi d'associer un poids nul aux larves mortes au cours de l'expérience. Comme le dispositif était un plan factoriel en randomisation totale pour chaque essai de chaque larve étudiée, nous avons choisi le modèle suivant linéaire mixte suivant :

$\forall i, l$, et m avec $i = 1, \dots, 8$; $l = 1, \dots, 8$ et $m = 1, \dots, 4$:

$$\boxed{y_{ilm} = \nu + \alpha_i + \delta_l + (\alpha\delta)_{il} + D_{im} + \varepsilon_{ilm}}$$

Où :

- y_{ilm} est le poids d'une larve de la répétition m , à la date l , subissant le traitement i ,
- ν est le paramètre de la moyenne générale,
- α_i est l'effet fixe du traitement i ,
- δ_l est l'effet fixe de la date l ,
- $(\alpha\delta)_{il}$ est l'effet fixe de l'interaction du traitement i , à la date l ,
- D_{im} est l'effet aléatoire de la répétition m du traitement i ,
- $\varepsilon_{ilm} \sim \mathcal{N}(0, \sigma^2 I)$ est l'erreur résiduelle de la répétition m , à la date l , pour le traitement i .

Nous allons maintenant aborder les points théoriques rattachés aux modèles utilisés.

Les modèles linéaires généralisés mixtes (ou GLMMs) sont une généralisation à la fois des modèles linéaires généralisés et des modèles linéaires mixtes. C'est pourquoi nous allons donc dans un premier temps rappeler la définition et les propriétés de ces deux types de modèles, avant d'aborder la partie théorique concernant le GLMM.

2.3 Modèles linéaires généralisés

2.3.1 Structure du modèle

Les modèles linéaires généralisés (GLMs) ont été introduits par Nelder et Wedderburn (1972) comme une extension des modèles linéaires classiques, et ont été formalisés par Mc Cullagh et Nelder (1989).

Ces modèles sont caractérisés par trois composantes : la distribution de y , le prédicteur linéaire, et la fonction de lien.

2.3.1.1 Distribution de y

On suppose que l'échantillon statistique y est constitué de n variables aléatoires $\{y_i ; i = 1, \dots, n\}$ indépendantes admettant des distributions appartenant à la famille exponentielle. Cela signifie que les lois de ces variables sont dominées par cette même mesure dite de référence et que la famille de leur densité par rapport à cette mesure s'écrit sous la forme :

$$f(y_i; \theta_i, \phi) = \exp \left\{ \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right\}$$

Où θ_i est appelé le paramètre naturel de la famille exponentielle, et $a(\phi)$ le paramètre de dispersion (McCulloch et Searle 2001 p 138 et 154) sur lequel nous reviendrons par la suite.

Remarque : dans certains cas $a(\phi)$ peut dépendre du type d'observation, mais ici on se limitera aux situations dans lesquelles $a(\phi)$ est homogène.

La famille exponentielle regroupe la plupart des lois usuelles comportant un ou deux paramètres : gaussienne, gaussienne inverse, gamma, poisson, binomiale ...

Ainsi, on en déduit la log-vraisemblance de la loi jointe de y_1, \dots, y_n :

$$l(\theta, \phi; y_1, \dots, y_n) = \sum_{i=1}^n \left\{ \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right\}$$

qui s'écrit également sous la forme matricielle suivante :

$$l(\theta, N; Y) = Y^t A^{-1} \theta - (b_\theta)^{1/2} A^{-1} (b_\theta)^{1/2} + \mathbb{1}^t c$$

où: A est la matrice diagonale dans laquelle $a(\phi)$ est le $i^{\text{ème}}$ élément de la diagonale,

θ est un vecteur dont le $i^{\text{ème}}$ élément est θ_i ,

b_θ est un vecteur dont le $i^{\text{ème}}$ élément est $b(\theta_i)$,

c est un vecteur dont le $i^{\text{ème}}$ élément est $c(y_i, \phi)$,

$\mathbb{1}$ est le vecteur unitaire de dimension n .

En annexe sont présentées les trois distributions utilisées dans ce rapport : gaussienne, poissonnienne, et binomiale, ainsi que certaines de leurs propriétés.

2.3.1.2 Prédicteur linéaire

Les observations planifiées des variables explicatives sont organisées dans la matrice X du plan.

Soit β un vecteur de p paramètres, le prédicteur linéaire est le vecteur à n composantes :

$$\eta = X\beta$$

Ce prédicteur est la composante déterministe du modèle.

2.3.1.3 Fonction de lien

Soit y_i une variable aléatoire dont la distribution appartient à la famille exponentielle tel que :

$$y_i = \mu_i + \varepsilon_i \text{ avec : } \mathbb{E}(y_i) = \mu_i \text{ et } \varepsilon_i \text{ suivant une loi d'espérance nulle.}$$

La troisième composante du modèle linéaire généralisé consiste à exprimer une relation fonctionnelle entre la composante aléatoire μ et le prédicteur linéaire :

$$g(\mu_i) = X_i^t \beta$$

Où g est la fonction de lien ; elle est monotone et différentiable (McCullagh et Nelder 1989 p27).

Remarque : la fonction de lien qui associe l'espérance μ_i au paramètre naturel est appelée la fonction de lien canonique, et dans ce cas :

$$g(\mu_i) = \theta_i = X_i^t \beta$$

L'inverse h de la fonction de lien est défini par : $h(\eta) = g^{-1}(\eta) = \mu$ permet d'obtenir une estimation de μ à l'aide de l'estimation $\hat{\beta}$ de β :

$$\hat{\mu} = g^{-1}(X\hat{\beta})$$

2.3.1.4 Expression des moments

L'étude du maximum de la log-vraisemblance nécessite la connaissance des dérivées suivantes:

$$\frac{\partial l}{\partial \theta_i} = \frac{y_i - b'(\theta_i)}{a(\phi)}$$

$$\frac{\partial^2 l}{\partial \theta_i^2} = \frac{-b''(\theta_i)}{a(\phi)}$$

Pour les lois issues de structure exponentielle, les conditions de régularité vérifiées permettent d'écrire :

$$\mathbb{E} \left(\frac{\partial l}{\partial \theta} \right) = 0 \quad \text{et} \quad - \mathbb{E} \left(\frac{\partial^2 l}{\partial \theta^2} \right) = \mathbb{E} \left(\frac{\partial l}{\partial \theta} \right)^2.$$

D'où :

$$\mathbb{E}(y_i) = \mu_i = b'(\theta_i)$$

De plus, comme :

$$\mathbb{E} \left(\frac{b''(\theta_i)}{a(\phi)} \right) = \mathbb{E} \left(\frac{y_i - b'(\theta_i)}{a(\phi)} \right)^2 = \frac{Var(y_i)}{a^2(\phi)}$$

on a donc :

$$\boxed{Var(y_i) = b''(\theta_i)a(\phi) = V(\mu_i)a(\phi)}$$

Où $V(\cdot)$ est la fonction de variance (McCullagh et Nelder 1989 p 28-29).

On peut décrire la structure de la matrice de variance-covariance des observations en fonction de ϕ et de V :

$$Var(Y) = R = R_{\mu}^{\frac{1}{2}} A R_{\mu}^{\frac{1}{2}}$$

Où R_{μ} est la matrice diagonale dans laquelle $V(\mu_i)$ est le $i^{\text{ème}}$ élément de la diagonale, et A telle que définie précédemment.

2.3.2 Estimation dans les modèles linéaires généralisés

2.3.2.1 Estimation par maximum de vraisemblance

On se trouve dans le cas d'un modèle linéaire généralisé défini par :

$$\eta_i = X_i^t \beta = g(\mu_i)$$

On suppose que les n observations sont indépendantes.

Soit $L(\beta) = \sum_{i=1}^n l_i(\theta_i, \phi; y_i)$ la log-vraisemblance.

On remarque (McCullagh et Nelder 1989 p 40-43) que :

$$\begin{aligned} \frac{\partial l_i}{\partial \beta_j} &= \frac{\partial l_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} \\ \frac{\partial l_i}{\partial \theta_i} &= \frac{y_i - b'(\theta_i)}{a(\phi)} = \frac{y_i - \mu_i}{a(\phi)} \\ \frac{\partial \mu_i}{\partial \theta_i} &= \frac{\partial b'(\theta_i)}{\partial \theta_i} = b''(\theta_i) = \frac{Var(y_i)}{a(\phi)} \end{aligned}$$

$$\frac{\partial \eta_i}{\partial \beta_j} = X_{ij} \quad \text{car } \eta_i = X_i^t \beta$$

Ainsi, les équations de maximisation de la log-vraisemblance s'écrivent :

$$\begin{aligned} \frac{\partial L(\beta)}{\partial \beta} &= 0 \\ \Leftrightarrow \sum_{i=1}^n \frac{\partial l_i(\theta_i, \phi; y_i)}{\partial \beta_j} &= 0 \\ \Leftrightarrow \boxed{\sum_{i=1}^n \frac{(y_i - \mu_i) X_{ij}}{\text{Var}(y_i)} \frac{\partial \mu_i}{\partial \eta_i}} &= 0 \quad j = 1, \dots, p \end{aligned}$$

Ce système d'équations peut s'écrire encore sous la forme :

$$X^t W_\beta^t \frac{\partial \eta}{\partial \mu} (y - \mu) = 0$$

avec $W_\beta = \text{diag} \{ \text{Var}(Y_i) g'(\mu_i)^2 \}_{i=1, \dots, n}$

$$\frac{\partial \eta}{\partial \mu} = \text{diag} \left\{ \frac{\partial \eta_i}{\partial \mu_i} \right\}_{i=1, \dots, n} = \text{diag} \{ g'(\mu_i) \}_{i=1, \dots, n}$$

Ces équations ne sont pas linéaires en β , c'est pourquoi leurs résolutions nécessitent des méthodes itératives dans lesquelles interviennent le Hessien (pour l'algorithme de Newton-Raphson) ou la matrice d'information (pour l'algorithme des scores de Fisher).

L'algorithme de Fisher peut aussi être décrit comme la résolution itérative des équations normales :

$$X^t W_\beta^{-1} (y^* - X\beta) = 0$$

où $y^* = \eta + \frac{\partial \eta}{\partial \mu} (y - \mu) = X\beta + \frac{\partial \eta}{\partial \mu} (y - \mu)$.

A β fixé, on retrouve alors les équations des moindres carrés généralisés associées au modèle :

$Y^* = X\beta + \varepsilon$, avec $E(\varepsilon) = 0$, et W_β la matrice de variance covariance de Y^* .

Le processus est itératif car Y^* et W dépendent des estimations en cours de η et de μ .

2.3.2.2 Estimation par maximum de quasi-vraisemblance

Dans certains problèmes, on ne peut pas expliciter la fonction de vraisemblance. McCulloch et Searle (2001 p 152-153) proposent d'utiliser uniquement les deux premiers moments du vecteur Y (à composantes indépendantes).

On définit la log-quasi-vraisemblance comme suit :

$$Q(\mu; y) = \sum_{i=1}^n \int_{y_i}^{\mu_i} \frac{y_i - t}{a(\phi) V(t)} dt$$

Pour trouver l'estimateur de β par maximum de quasi-vraisemblance (MQL), on doit résoudre l'équation suivante :

$$\begin{aligned} \frac{\partial Q(\mu; y)}{\partial \beta} &= 0 \\ \Leftrightarrow \sum_{i=1}^n \frac{y_i - \mu_i}{a(\phi) V(\mu_i)} \frac{\partial \mu_i}{\partial \beta} &= 0 \\ \Leftrightarrow \sum_{i=1}^n \frac{y_i - \mu_i}{a(\phi) V(\mu_i)} \frac{X_i^t}{g'(\mu_i)} &= 0 \end{aligned} \quad (1)$$

$$\text{car : } \frac{\partial \mu_i}{\partial \beta} = \frac{\partial \mu_i}{\partial g(\mu_i)} \frac{\partial g(\mu_i)}{\partial \beta} = \left(\frac{\partial g(\mu_i)}{\partial \mu_i} \right)^{-1} X_i^t = \frac{X_i^t}{g'(\mu_i)}$$

Le système d'équations (1), peut s'écrire sous la forme matricielle suivante :

$$X^t W_\beta^{-1} \frac{\partial \eta}{\partial \mu} (y - \mu) = 0$$

qui est une procédure itérative.

$\frac{\partial \eta}{\partial \mu}$ a été défini précédemment, et on rappelle que W_β est défini par :

$$W_\beta = \text{diag} \{ \text{Var}(Y_i) g'(\mu_i)^2 \}_{i=1, \dots, n} = \text{diag} \{ a(\phi) V(\mu_i) g'(\mu_i)^2 \}_{i=1, \dots, n} .$$

On retrouve bien le même système d'équations que pour le maximum de vraisemblance décrit précédemment. L'intérêt de cette méthode est que l'on peut envisager des fonctions de variance différentes de celles associées aux lois classiques.

Dans le cas de la distribution de Poisson, par exemple, la méthode par maximum de vraisemblance supposera que $\text{Var}(y_i) = V(\mu_i)$, ce qui en pratique, n'est pas souvent vérifié. En revanche, si la variance est proportionnelle à l'espérance, le modèle estimé par maximum de quasi-vraisemblance sera toujours correct car il suppose seulement que : $\text{Var}(y_i) = a(\phi) V(\mu_i)$.

2.4 Modèles linéaires mixtes

2.4.1 Description du modèle

Le modèle linéaire mixte (ou modèle linéaire à effets aléatoires) est défini par :

$$Y = X\beta + ZU + \varepsilon = X\beta + \sum_{k=1}^K Z_k U_k + \varepsilon$$

$X\beta$ = partie effets fixes

ZU = partie effets aléatoires

ε = variable aléatoire résiduelle

Où :

$\varepsilon \sim \mathcal{N}_{\mathbb{R}^N}(0, R)$ avec $R = \theta_0 V_0$,

$U_k = \mathcal{N}_{\mathbb{R}^{q_k}}(0, \theta_k G_k)$ pour tout $k = 1, \dots, K$.

V_0 et les G_k sont des matrices connues.

U_1, \dots, U_k sont k effets aléatoires non observés, **indépendants entre eux** et **indépendants de ε** .

On note $G_\theta = \text{diag}\{\theta_k G_k\}_{k=1, \dots, K}$, et $U \sim \mathcal{N}_{\mathbb{R}^q}(0, G_\theta)$ où $q = \sum_{k=1}^K q_k$.

(McCulloch et Searle 2001 p 157-161).

2.4.2 Moments conditionnels et marginaux

- Moments conditionnels aux effets aléatoires :

$$\mathbb{E}(Y | U) = X\beta + ZU$$

$$\text{Var}(Y | U) = R$$

- Moments marginaux :

$$\mathbb{E}(Y) = X\beta$$

$$\text{Var}(Y) = \Gamma_\theta = R + ZG_\theta Z^t = \sum_{k=0}^K \theta_k V_k \quad \text{où } V_k = Z_k G_k Z_k^t.$$

Les $\theta_k \in \mathbb{R}^+$ sont appelés « composantes de la variance ».

2.4.3 Autres propriétés

2.4.3 Autres propriétés

$$\begin{pmatrix} Y \\ U \end{pmatrix} \sim \mathcal{N}_{N+q} \left(\begin{pmatrix} X\beta \\ 0 \end{pmatrix}, \begin{pmatrix} R + ZG_\theta Z^t & ZG_\theta \\ G_\theta Z^t & G_\theta \end{pmatrix} \right)$$

$$U | Y \sim \mathcal{N}_q (G_\theta Z^t \Gamma_\theta^{-1} (y - X\beta), G_\theta - G_\theta Z^t \Gamma_\theta^{-1} Z G_\theta)$$

2.4.4 Vraisemblance

La fonction de vraisemblance s'écrit :

$$f(\beta, \theta; y) = \frac{1}{(2\pi)^{\frac{N}{2}} |\Gamma|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (y - X\beta)^t \Gamma^{-1} (y - X\beta) \right\}$$

et la log-vraisemblance :

$$l(\beta, \theta; y) = -\frac{N}{2} \ln(2\pi) - \frac{1}{2} \ln(|\Gamma|) - \frac{1}{2} (y - X\beta)^t \Gamma^{-1} (y - X\beta)$$

2.4.5 Estimation dans les modèles linéaires généralisés mixtes

2.4.5.1 Estimation par maximum de vraisemblance

Les estimations de β et de θ sont solutions des équations suivantes :

$$\begin{cases} \frac{\partial l(\beta, \theta; y)}{\partial \beta} = 0 \\ \frac{\partial l(\beta, \theta; y)}{\partial \theta_j} = 0 \quad \forall j \in \{0, \dots, K\} \end{cases}$$

D'après l'expression de la log-vraisemblance décrite ci-dessus, on obtient :

$$\frac{\partial l(\beta, \theta; y)}{\partial \beta} = X^t \Gamma^{-1} (y - X\beta) = 0$$

Et, comme on a les propriétés suivantes (Searle et al. 1992), quelque soit A une matrice:

$$\frac{\partial \log |A|}{\partial \theta_j} = \text{tr} \left(A^{-1} \frac{\partial A}{\partial \theta_j} \right) \quad \text{et} \quad \frac{\partial A^{-1}}{\partial \theta_j} = -A^{-1} \frac{\partial A}{\partial \theta_j} A^{-1}$$

On a donc pour tout $j \in \{0, \dots, K\}$:

$$\frac{\partial l(\beta, \theta; y)}{\partial \theta_j} = -\frac{1}{2} \text{tr}(\Gamma^{-1} V_j) + \frac{1}{2} (y - X\beta)^t \Gamma^{-1} V_j \Gamma^{-1} (y - X\beta) = 0$$

On obtient ainsi le système d'équations suivant (Trottier 1998 p41-43):

$$\begin{cases} X^t \Gamma^{-1} X \beta = X^t \Gamma^{-1} y \\ \text{tr}(\Gamma^{-1} V_j) = y^t P V_j P y \quad \forall j \in \{0, \dots, K\} \end{cases} \quad (2)$$

Où $P = \Gamma^{-1}(I - X(X^t \Gamma^{-1} X)^{-1} X^t \Gamma^{-1})$.

Remarque : si une des composantes de la variance est connue, on reporte dans Γ sa valeur, sinon on la remplace par son estimation.

Par ailleurs, les $K + 1$ équations, concernant uniquement les composantes de la variance, ne sont pas linéaires en θ . Il est donc nécessaire d'utiliser une méthode itérative pour les résoudre.

La transformation suivante :

$$\text{tr}(\Gamma^{-1} V_j) = \text{tr}(\Gamma^{-1} V_i \Gamma^{-1} \Gamma) = \sum_{j=0}^K \theta_j \text{tr}(\Gamma^{-1} V_i \Gamma^{-1} V_j)$$

permet de mettre en place l'algorithme itératif :

$$\begin{cases} X^t \Gamma^{-1} X \beta = X^t \Gamma^{-1} y \\ (\text{tr}(\Gamma^{-1} V_i \Gamma^{-1} V_j))_{i,j=0,\dots,K} = (y^t P V_j P y)_{j=0,\dots,K} \end{cases} \quad (3)$$

On peut donc dans un premier temps résoudre les équations (3), à partir des valeurs initiales de θ , puis, en fonction des valeurs obtenues, obtenir une estimation de β .

Cependant, ces nouvelles équations ne sont toujours pas linéaires en θ et l'inversion de Γ reste difficile.

Remarque : l'estimation par maximum de vraisemblance restreint équivaut à une estimation par maximum de vraisemblance dans le projeté du modèle étudié. L'estimation est également obtenue par les équations (2) en remplaçant seulement Γ^{-1} par P .

L'estimation des effets fixes et des composantes de la variance peut aussi être obtenue par une autre méthode que nous allons décrire maintenant.

2.4.5.2 Equations de Henderson

Ces équations (Henderson, Kempthorne, Searle, et VonKrosig 1959) permettent d'obtenir la prédiction BLUP (Best Linear Unbiased Predictor) de U et l'estimation BLUE (Best Linear Unbiased Estimator) de β .

- Equations du modèle mixte (MME)

La log-vraisemblance jointe (Y, U) est définie par :

$$l(\beta, U; y, U) = -\frac{N+q}{2} \ln(2\pi) - \frac{1}{2} \ln(|R|) - \frac{1}{2} \ln(|G_\theta|) - \frac{1}{2} \{ (y - X\beta - ZU)^t R^{-1} (y - X\beta - ZU) + U^t G_\theta^{-1} U \}$$

L'approche d'Henderson consiste à maximiser $l(\beta, U; y, U)$ par rapport à β et U . Le résultat de cette maximisation se traduit par le système suivant:

$$\begin{pmatrix} X^t R^{-1} X & X^t R^{-1} Z \\ Z^t R^{-1} X & Z^t R^{-1} Z + G_\theta^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ U \end{pmatrix} = \begin{pmatrix} X^t R^{-1} y \\ Z^t R^{-1} y \end{pmatrix}$$

Ces équations sont appelées « **équations du modèle mixte** » (MME).

Remarque : sans la présence de G_θ^{-1} , on retrouve les équations d'un modèle linéaire (ML) à effet fixe pour U .

- Solutions

Les solutions du système encadré ci-dessus sont :

$$\hat{\beta} = BLUE(\beta) = (X^t \Gamma_\theta^{-1} X)^{-1} X^t \Gamma_\theta^{-1} y$$

$$\hat{U} = BLUP(U) = G_\theta Z^t \Gamma_\theta^{-1} (y - X\hat{\beta})$$

Les équations MME sont plus économiques en calculs que les équations ML. Les équations ML requièrent une inversion de la matrice Γ_θ d'ordre N (= nombre d'observations). Alors que les équations MME nécessitent seulement de l'inversion d'une matrice d'ordre $p+q$ qui correspond au nombre total de niveaux des effets fixes et des effets aléatoires, et qui est en général bien inférieur à N .

Les équations du modèle mixte requièrent l'inversion des matrices R et G_θ . Ces matrices étant toujours diagonales, l'inversion est par conséquent aisée.

- Estimation des composantes de la variance

Les solutions du système d'Henderson décrit ci-dessus permettent d'obtenir les estimations ML et REML des composantes de la variance par le biais de schémas itératifs (Trottier 1998 p 53-55).

2.5 Modèles linéaires généralisés mixtes

2.5.1 Description du modèle

Les modèles linéaires généralisés mixtes sont une généralisation des deux types de modèle que l'on vient de décrire précédemment (McCulloch et Searle 2001 p 221-222).

Dans un tel modèle on suppose ainsi que :

- Les composantes de y , conditionnellement à U , sont indépendantes et de loi appartenant à la famille exponentielle.
- De plus, le prédicteur linéaire est défini tel que :

$$g(\mu) = X\beta + ZU$$

où :

- la fonction de lien g est monotone et différentiable,
- $U \sim \mathcal{N}(0, G_\theta)$

Les autres éléments ont été définis précédemment.

Les moments conditionnels sont définis comme suit:

$$\mathbb{E}(Y | U) = \mu = g^{-1}(X\beta + ZU)$$

$$\forall i \in \{1, \dots, n\} : \text{Var}(y_i | U) = V(\mu_i) a(\phi) \quad \text{et} \quad \text{Var}(Y | U) = \text{diag} \{V(\mu_i) a(\phi)\}_{i \in \{1, \dots, n\}}$$

(Trottier 1998 p 26).

La vraisemblance jointe de (y, U) est obtenue en multipliant la densité conditionnelle de $y|U$ par la densité marginale de U (cf : la log-vraisemblance jointe décrite dans la section précédente).

2.5.2 Estimations dans les GLMMs

2.5.2.1 Plusieurs méthodes d'estimation

La difficulté majeure dans l'estimation des modèles linéaires généralisés mixtes est la présence d'intégrales sans solutions analytiques (excepté dans des cas spécifiques) dans la vraisemblance jointe de (y, U) .

Les principales stratégies utilisées pour traiter ces intégrales sont :

- Intégration numérique avec la quadrature de Gauss-Hermite,
- Intégration numérique avec des méthodes de Monte-Carlo,
- Approximation de l'intégrale avec la méthode de Laplace,
- Eviter l'intégration en linéarisant le modèle par le biais des séries de Taylor.

La méthode PQL (Penalized quasi-likelihood) décrite par Breslow et Clayton (1993) peut-être obtenue avec une version modifiée de l'approximation de Laplace, ou en linéarisant le modèle.

Nous choisissons ici de présenter cette méthode tout d'abord car elle fait partie des méthodes d'estimation employées par la procédure GLIMMIX de SAS que nous décrirons par la suite. De plus, la méthode PQL découle de ce qui a été présenté précédemment puisque la première étape consiste à définir une fonction de quasi-vraisemblance (similaire à ce qui a été réalisé avec les GLM), et la deuxième étape utilise la même méthode d'estimation que les modèles linéaires mixtes standards.

Bien entendu il existe d'autres méthodes :

En dehors de la méthode PQL, Breslow et Clayton (1993) utilisent la méthode de quasi-vraisemblance marginale (MQL) dans laquelle la quasi-vraisemblance est définie par une approximation des deux premiers moments.

La méthode de Gilmour, Anderson et Rae (1985) ressemble à la méthode MQL mais avec une approximation sensiblement différente.

Schall (1991) mais aussi Engel et Keen (1992), proposent également des méthodes d'estimations dans les GLMMs, en procédant cette fois-ci en deux étapes : linéarisation, puis utilisation des équations de Henderson pour obtenir les estimations dans le modèle linéaire mixte associé.

2.5.2.2 Méthode de quasi-vraisemblance pénalisée (PQL)

Dans la suite, on considère que le paramètre de dispersion ϕ est fixé à 1. Cependant, dans d'autres applications il peut être estimé en même temps que θ comme un paramètre de la matrice de variance-covariance de la distribution marginale de y .

La fonction de quasi-vraisemblance intégrée (Breslow et Clayton 1993) est définie par:

$$\exp \{ ql(\beta, \theta) \} \propto |G_\theta|^{-\frac{1}{2}} \int \exp \left[-\frac{1}{2\phi} \sum_{i=1}^n d(y_i, \mu_i) - \frac{1}{2} U^t G_\theta^{-1} U \right] dU \quad (4)$$

où

$$d(y, \mu) = -2 \int_y^\mu \frac{y-t}{a(\phi) v(t)} dt \text{ est la mesure de déviance de l'ajustement.}$$

Remarque : dans le cas où $Y|U$ suit une loi appartenant à la famille exponentielle, on retrouve alors la formule de déviance suivante :

$$2 a(\phi) \{ l(y; y, \phi) - l(y; \mu, \phi) \}$$

- Critère PQL

L'objectif ici est d'estimer à la fois les coefficients des effets fixes β et ceux des effets aléatoires U (où $U \sim \mathcal{N}(0, G_\theta)$), en maximisant la quasi-log vraisemblance décrite ci-dessus.

En écrivant l'équation (4) sous la forme :

$$c |G_\theta|^{-\frac{1}{2}} \int \exp[-k(U)] dU \quad (5)$$

on peut appliquer la méthode de Laplace pour obtenir une approximation de l'intégrale.

Soit k' et k'' les dérivées premières de k par rapport à U .

En ne tenant pas compte de la constante multiplicative c , on parvient à l'approximation suivante :

$$ql(\beta, \theta) \approx -\frac{1}{2} \log |G_\theta| - \frac{1}{2} \log |k''(\tilde{U})| - k(\tilde{U})$$

où $\tilde{U} = \tilde{U}(\beta, \theta)$ est solution de :

$$k'(U) = -\sum_{i=1}^n \frac{(y_i - \mu_i) Z_i}{a(\phi) V(\mu_i) g'(\mu_i)} + G_\theta^{-1} U = 0$$

qui minimise $k(U)$, et permet ainsi de maximiser (**).

Par ailleurs, on a :

$$k''(U) = \sum_{i=1}^n \frac{Z_i Z_i^t}{a(\phi) V(\mu_i) [g'(\mu_i)]^2} + G_\theta^{-1} + C \quad (6)$$

$$\approx Z^t W^{-1} Z + G_\theta^{-1}$$

où W est définie comme précédemment : $W = \text{diag} \{ a(\phi) V(\mu_i) g'(\mu_i)^2 \}_{i=1, \dots, n}$

et : $C = -\sum_{i=1}^n (y_i - \mu_i) Z_i \frac{\partial}{\partial U} \frac{1}{a(\phi) V(\mu_i) g'(\mu_i)}$ est négligeable car d'ordre inférieur à celui des deux autres termes de l'expression.

En combinant (4) et (6), et en négligeant C , on obtient :

$$ql(\beta, \theta) \approx -\frac{1}{2} \log |I + Z^t W^{-1} Z G_\theta| - \frac{1}{2\phi} \sum_{i=1}^n d(y_i, \mu_i) - \frac{1}{2} \tilde{U}^t G_\theta^{-1} \tilde{U}$$

où \tilde{U} est choisi de façon à maximiser la somme des deux derniers termes.

En supposant que les poids du modèle linéaire généralisé varient lentement, on ignore le premier terme de l'expression ci-dessus et on choisit β de façon à ce qu'il maximise le second terme.

Ainsi, $(\hat{\beta}, \hat{U})$ où $\hat{U}(\theta) = \tilde{U}(\hat{\beta}(\theta))$ maximisent la quasi-vraisemblance pénalisée (PQL) de Green :

$$-\frac{1}{2\phi} \sum_{i=1}^n d(y_i, \mu_i) - \frac{1}{2} U^t G_\theta^{-1} U$$

En différenciant par rapport aux paramètres β et U , on arrive aux équations suivantes :

$$\sum_{i=1}^n \frac{y_i - \mu_i}{a(\phi) V(\mu_i)} \frac{X_i^t}{g'(\mu_i)} = 0 \quad (7)$$

On remarque qu'il s'agit bien ici de la même égalité obtenue avec les modèles GLM.

$$\sum_{i=1}^n \frac{y_i - \mu_i}{a(\phi) V(\mu_i)} \frac{Z_i^t}{g'(\mu_i)} = G_\theta^{-1} U \quad (8)$$

- *Algorithme des scores de Fisher (Fisher scoring)*

Green (1987) a développé l'algorithme de Fisher afin de pouvoir résoudre les équations (7) et (8) comme un problème itératif des moindres carrés pondérés.

Soit y^* tel que : $y_i^* = \eta_i + (y_i - \mu_i) g'(\mu_i)$, et W la matrice de poids telle que définie précédemment.

D'après l'algorithme de Fisher, les solutions des équations (7) et (8) s'écrivent sous la forme du système suivant :

$$\begin{pmatrix} X^t W^{-1} X & X^t W^{-1} Z \\ Z^t W^{-1} X & Z^t W^{-1} Z + G_\theta^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ U \end{pmatrix} = \begin{pmatrix} X^t W^{-1} y \\ Z^t W^{-1} y \end{pmatrix}$$

où $U = G_\theta V$.

On constate que l'on retrouve ici les équations de Henderson. Par conséquent, comme dans la partie sur l'estimation dans les modèles linéaires mixtes classiques, on peut obtenir les estimations de β et U de qui suivent :

$$\hat{\beta} = BLUE(\beta) = (X^t \Gamma_\theta^{-1} X)^{-1} X^t \Gamma_\theta^{-1} y$$

$$\hat{U} = BLUP(U) = G_\theta Z^t \Gamma_\theta^{-1} (y - X\hat{\beta})$$

2.6 Surdispersion

2.6.1 Mise en évidence de la surdispersion

Parfois, il arrive que la variance des erreurs ne coïncide pas avec celle que prévoit la théorie. Par exemple, pour le modèle de Poisson, en théorie $\mathbb{E}(Y) = \text{Var}(Y)$ mais en pratique la variance est souvent beaucoup plus élevée que la moyenne, et on parle alors de surdispersion.

On peut mettre en évidence la présence de surdispersion dans les données en estimant le paramètre de surdispersion $a(\phi)$ (ou paramètre d'échelle). Une estimation de ce paramètre est donnée par :

$$\widehat{a(\phi)} = \frac{\chi^2}{n - p}$$

où n est le nombre d'observation, et p est le nombre de paramètres dans le modèle, et où χ^2

est le Khi-deux généralisé de Pearson : $\chi^2 = \sum \frac{(y_i - \widehat{\mu}_i)^2}{V(\widehat{\mu}_i)}$, ($\widehat{\mu}$ l'estimateur du maximum de vraisemblance de μ et V la fonction de variance. (McCullagh et Nelder, 1989, p200)).

Soit la d déviance, alors si le modèle ajuste bien les données (Dobson 1990 p 60) on a également :

$$\widehat{a(\phi)} = \frac{d}{n - p}$$

Par ailleurs, Dean (1992) propose une statistique de test permettant d'évaluer la surdispersion liée à un modèle de Poisson qui est basé sur le test du score :

H_0 : modèle de Poisson sans surdispersion

H_1 : modèle de Poisson avec surdispersion

La statistique de test est définie par :

$$T = \frac{\sum_i [(y_i - \widehat{\mu}_i)^2 - \widehat{\mu}_i]}{\sqrt{2 \sum_i \widehat{\mu}_i^2}}$$

Sous l'hypothèse nulle, cette statistique suit une loi $\mathcal{N}(0,1)$.

L'idée de ce test repose sur le fait que sous H_0 , \widehat{y}_i et $\frac{\sum_i (y_i - \mu_i)^2}{n}$ (le numérateur de T évalue la différence entre ces deux quantités), estiment la même quantité alors que sous l'hypothèse H_1 , le second terme est d'autant plus grand que la surdispersion est importante (Marque 2003 p 29-30).

Enfin, un examen graphique peut aussi contribuer à la mise en évidence d'une éventuelle surdispersion. Dans le cas du modèle de Poisson, on peut distinguer la surdispersion en traçant le graphe des variances empiriques en fonction des moyennes, et la droite de régression. Si la droite de régression a un coefficient

directeur plus élevé que la droite d'équation $y = x$, alors on peut supposé qu'il y a surdispersion (Allain et Brenac 2001).

2.6.2 Paramètre de dispersion

Le paramètre de dispersion permet d'augmenter ou de diminuer la variance estimée par la distribution.

On suppose que $a(\phi) > 0$.

Si $a(\phi) = 1$ alors il n'y a ni sous dispersion, ni surdispersion.

Si $a(\phi) < 1$ de façon significative, alors il y a sous-dispersion des données, causant une sur-estimation des erreurs standards des paramètres et une sous-estimation des statistiques de tests.

Si $a(\phi) > 1$ de façon significative, alors on est dans le cas d'une surdispersion avec comme conséquences un gonflement irréaliste des statistiques de tests et une sous-estimation des erreurs standards des paramètres.

Dans les deux derniers cas, un ajustement est nécessaire.

Pour prendre en compte la surdispersion, on utilise la formule qui suit:

$$Var(y) = V(\mu)a(\phi)$$

2.6.3 Conséquences de l'ajustement

Ajouter un paramètre de dispersion à la fonction de variance peut avoir plusieurs conséquences.

Tout d'abord, la surdispersion est étroitement liée à la distribution étudiée. En effet, des données étant surdispersées pour la loi de poisson ne le seront pas nécessairement pour la loi binomiale négative (Littell, et al. 2006 p 540-541).

De plus, il existe plusieurs causes différentes de surdispersion : la corrélation entre certaines données, les éventuelles erreurs dans l'écriture du modèle dues à l'omission d'une (ou plusieurs) variable, ou un mauvais choix de distribution ou de fonction de lien...

Cas particuliers des effets aléatoires :

Par ailleurs, les modèles linéaires mixtes peuvent en un sens modéliser la surdispersion (McCulloch, et Searle 2001 p 222-223). En effet, dans le cas par exemple d'un modèle de Poisson, l'ajout d'effets aléatoires a la conséquence suivante sur la distribution marginale :

$$Var(Y) > E(Y)$$

Ainsi, on peut en déduire que les effets aléatoires constituent ici un moyen de modéliser ou d'attribuer de la surdispersion à une distribution marginale.

Abordons maintenant la description des outils « techniques » nécessaires à ce projet.

2.7 Logiciels

Nous exposons brièvement ici les procédures utilisées dans cette étude, afin de les comparer.

2.7.1 Fonction glmer de R

La fonction `glmer` du package `lme4` de R permet d'ajuster des modèles linéaires généralisés mixtes. La syntaxe simplifiée de cette fonction que nous avons utilisée est la suivante :

```
glmer ( formule du modèle, family= distribution choisie, data= table de données)
```

Par exemple, pour ajuster le GLMM permettant de modéliser le nombre de vers blancs (cf 3.2), on utilise :

```
modele = glmer (nb_vers ~ bloc + traitement + horizon + traitement : horizon + (1 | bloc : traitement),  
family = quasipoisson, data = densite_vb)
```

où `nb_vers` est la variable à expliquer, `traitement : horizon` représente l'effet de l'interaction entre le facteur traitement et le facteur horizon, et `(1 | bloc : traitement)` représente l'effet aléatoire de l'interaction entre le facteur bloc et le facteur traitement. On note que la surdispersion est prise en compte en spécifiant « quasi » devant la distribution choisie.

L'estimation des effets fixes et aléatoires se fait par la méthode PQL que nous avons décrite précédemment.

Cette fonction crée un modèle de type « mer » sur lequel on peut appliquer différentes fonctions. Notamment, la fonction `summary()` renvoie les statistiques d'ajustement, l'estimation de la matrice de variance-covariance, l'estimation des effets fixes, ainsi que la matrice de corrélation des effets fixes. La fonction `residual()` permet d'accéder aux résidus. La fonction `fitted()` permet d'obtenir les valeurs prédites.

2.7.2 Procédure GLIMMIX de SAS

La procédure GLIMMIX de SAS permet également d'ajuster des modèles GLMMs. Pour l'exemple décrit avec `glmer`, la syntaxe sous SAS est la suivante :

```
proc glimmix ic=pq data=densite_VB;  
  class bloc parcelle T H;  
  model VersBlancs = bloc traitement | horizon / link=log s dist=poisson;  
  random bloc*traitement / subject=parcelle;  
  random _residual_;  
  lsmeans traitement*horizon / pdiff slice=horizon;  
  output out=TABLE pred=valeurs_predites resid=residus;  
run;  
quit;
```

où :

- `ic = pq` permet d'afficher les statistiques d'ajustement (AIC, BIC...),
- `class` spécifie les facteurs du modèle
- `model` décrit le modèle sous la forme : variable à expliquer = liste des effets fixes (l'interaction de deux facteurs A et B s'écrit $A*B$, et $A|B$ équivaut à $A + B + A*B$)
- `random` spécifie les effets aléatoires
- `lsmeans` renvoie les estimations des moyennes par la méthode des moindres carrés; en spécifiant `pdiff`, on obtient des comparaisons de moyennes deux à deux
- la dernière ligne crée une table contenant les valeurs prédites et les résidus

La procédure décrite ci-dessus renvoie des statistiques d'ajustements, l'estimation de la matrice de variance-covariance, l'estimation des effets fixes, les résultats d'une ANOVA de type III, les estimations des moyennes, les comparaisons de moyennes deux à deux par la méthode des contrastes.

GLIMMIX utilise une méthode par quasi-vraisemblance pour effectuer ces estimations (Littell, et al. 2006). Plusieurs méthodes d'estimations sont possibles : les approches PL et REPL de Wolfinger et O'Connell (1993), ou PQL et MQL de Breslow et Clayton (1993).

2.7.3 Comparaison

Deux points majeurs différencient principalement ces deux fonctions : l'ANOVA, et l'étude des contrastes.

Tout d'abord, l'analyse de variance ne peut pas être réalisée sur des modèles créés par `glmer`. En effet, la fonction `anova` de R peut s'appliquer à des modèles linéaires mixtes (`lmer`), mais pas des GLMMs. Soit le modèle M défini avec `glmer`, alors `anova(M)` renvoie une l'erreur suivante:

"Single argument anova for GLMMs not yet implemented."

Cette fonction `anova` peut en revanche prendre en argument deux modèles emboîtés $M1$ et $M2$ de type `glmer` : `anova(M1,M2)`. Et dans ce cas, elle effectue alors un test de rapport de vraisemblance.

Avec la procédure GLIMMIX, l'analyse de variance ne pose aucune difficulté.

De plus, il semblerait également qu'on ne puisse pas effectuer dans R des comparaisons de moyennes deux à deux par la méthode des contrastes, pour des modèles de type `glmer`. Alors, que c'est possible avec GLIMMIX.

Par ailleurs, il faut préciser ici certains points concernant les estimations des paramètres. L'estimation du paramètre de surdispersion renvoyée par `glmer` correspond à $\widehat{\alpha(\phi)}$, tandis que l'estimation donnée par GLIMMIX est en réalité égale à $\sqrt{\widehat{\alpha(\phi)}}$. De même, les estimations des moyennes réalisées avec la fonction `fitted` de R correspondent μ à alors qu'avec SAS, il s'agit en fait de $g(\mu)$ où g est la fonction de lien utilisée dans le GLMM considéré.

3. Résultats et discussion

Cette partie est consacrée tout d'abord à la description du protocole mis en place. Puis, nous détaillons l'application des méthodes présentées précédemment sur une variable en particulier : celle des densités de vers blancs dans les cultures de riz au mois de mars 2009, avant de procéder à une présentation synthétique de tous les résultats obtenus. Enfin, nous tentons d'évaluer les limites de ce travail, et de proposer des solutions pour les travaux à venir.

3.1 Protocole suivi pour chaque variable

Pour débiter les analyses statistiques, nous avons tout d'abord utilisé le logiciel R. Mais, face aux difficultés rencontrées, nous avons dû nous tourner vers une autre procédure : la « proc glimmix » de SAS.

L'étude des variables citées précédemment s'est effectuée en plusieurs étapes.

Comme dans toute analyse statistique, nous avons tout d'abord débuté par une analyse exploratoire des données. De plus, lorsque les distributions envisagées étaient poissonniennes ou binomiales, une étude de la dispersion des données a été réalisée.

L'ajustement du modèle a été effectué à l'aide des deux fonctions décrites brièvement dans la partie matériel, et l'étape de validation du modèle a été réalisée en analysant les résidus.

Par la suite, nous avons analysé les effets fixes du modèle afin de tenter de mettre en évidence l'influence des différents traitements sur la quantité de vers blancs présents dans le sol.

Enfin, les moyennes ont été estimées, et lorsque cela s'avérait pertinent, elles ont été comparées deux à deux à l'aide de la méthode des contrastes.

Par ailleurs, nous avons étudié aussi les corrélations qu'il pouvait exister entre les densités de vers blancs et les densités des autres macrofaunes, ainsi qu'entre les densités de vers blancs et les indices de diversité.

3.2 Application des méthodes présentées aux densités de vers blancs dans le dispositif riz pluvial au mois de mars

3.2.1 Analyse exploratoire

- *statistiques descriptives* :

L'analyse exploratoire des données a été principalement réalisée à l'aide de la procédure *proc univariate*² de SAS qui regroupe les résultats essentiels dans le cadre d'une étude univariée de variables quantitatives (cf. *tableau 2*) : moments, test de tendance centrale, test de normalité...

Avec une médiane de 1, on en déduit que les densités de vers blancs étaient très faibles, et cela s'observe aussi sur l'histogramme (cf *graphe 1*).

Par ailleurs, en étudiant indépendamment chaque horizon, on a pu constaté que la litière contenait 75% de valeurs nulles, et l'horizon 20-30cm près de 50%, contre seulement 4% et 8 % pour les profondeurs 0-10 et 10-20 cm.

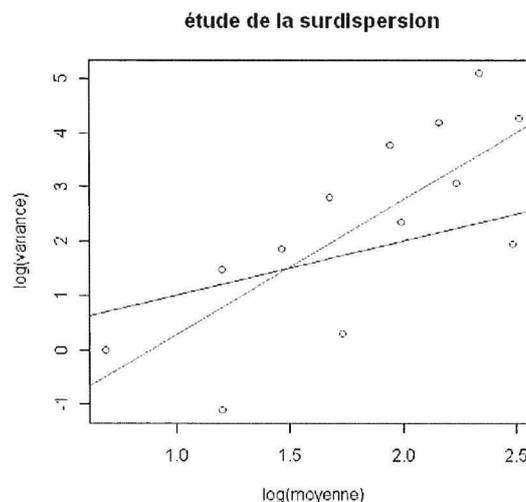
C'est pourquoi nous avons fait le choix de traiter uniquement les données des horizons 0-10 et 10-20 cm. Dans toute la suite de ce rapport, seuls ces deux horizons seront donc considérés.

Les box plot obtenus avec la procédure *proc boxplot* de SAS (cf *graphes 2*), de la densité de vers blancs en fonction des traitements, des blocs et des horizons permettaient d'émettre certaines hypothèses : en particulier, on voit que le nombre moyen de vers blanc variait beaucoup d'un horizon à l'autre, et l'on peut donc supposer ici que l'effet de l'horizon était significatif. En revanche, il est difficile d'émettre des suppositions concernant les effets bloc et traitement.

- *analyse graphique de la surdispersion :*

La moyenne des densités de vers blancs des horizons 0 à 20 cm était de 6,02 , tandis que la variance, bien plus élevée était de 30,74. Ainsi, on peut supposer que l'on se trouvait dans le cas de données surdispersées.

Pour tenter de vérifier cette hypothèse, nous avons calculé les moyennes et les variances par traitement et horizon, et représenté graphiquement le log des variances obtenues en fonction du log des moyennes.



En vert est représentée la droite de régression, en rouge la droite d'équation $y=x$.

On observe donc que :

² Remarque : la fonction *summary* dans R permet également d'obtenir un résumé (plus restreint) des statistiques descriptives.

$\log(\text{Var}(y_{ij})) > \log(\overline{y_{ij}})$ où $\overline{y_{ij}}$ sont les moyennes par traitement i et horizon j .

On peut donc supposer qu'on est bien dans un cas de surdispersion. Nous allons tenter de le vérifier dans la section suivante.

3.2.2 Présentation du modèle

D'après la section 2.2.1 et l'analyse exploratoire réalisée ci-dessus, on a choisi de modéliser les densités de vers blancs dans les cultures de riz en mars (pour les horizons 0-10 et 10-20 cm), par le GLMM avec une distribution quasi-poissonienne suivant :

$$y = \mu + \varepsilon$$

$$\log(\mu) = X\beta + ZU$$

avec $\forall i, j, k$ avec $i = 1, \dots, 8$; $j = 1, 2$ et $k = 1, \dots, 3$:

$$\log(\mu_{ijk}) = \nu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + B_{ik}$$

Où, on le rappelle:

- $y_{ijk} = \mu_{ijk} + \varepsilon_{ijk}$ est la densité de vers blancs de la « petite parcelle » i, j, k
- ν est le paramètre de la moyenne générale
- α_i est l'effet fixe du traitement i
- β_j est l'effet fixe de l'horizon j
- γ_k est l'effet fixe du bloc k
- $(\alpha\beta)_{ij}$ est l'effet fixe de l'interaction du traitement i avec l'horizon j
- B_{ik} est l'effet aléatoire de l'interaction du traitement i avec le bloc k
- ε_{ijk} est l'erreur résiduelle de la « petite parcelle » i, j, k .

Par ailleurs,

- X est une matrice fixée par l'expérience de dimension $n \times p$ avec le nombre d'observation $n = 48$, et $p = 18$.
- $\beta = (\nu, \alpha_1, \alpha_2, \dots, \alpha_7, \beta_1, \gamma_1, \gamma_2, (\alpha\beta)_{11}, (\alpha\beta)_{21}, \dots, (\alpha\beta)_{71})^t$ est le vecteur des paramètres des effets fixes de dimension $p \times 1$.
- Z est une matrice fixée par l'expérience de dimension $n \times q$, où $q = 18$.
- $U = (B_{11}, \dots, B_{ik}, \dots)_{\substack{i=1, \dots, 8 \\ j=1, 2}}^t \sim \mathcal{N}(0, G_\theta)$ avec $G_\theta = \sigma_{ik}^2 I$

Et comme la distribution est quasi-poissonienne, on a :

$$\mathbb{E}(y_{ijk} | B_{ik}) = \mu_{ijk}$$

$$\text{Var}(y_{ijk} | B_{ik}) = V(\mu_{ijk}) a(\phi) = \mu_{ijk} a(\phi)$$

$$= \text{Var}(\varepsilon_{ijk}) = \sigma_e^2$$

où $a(\phi)$ est le paramètre de dispersion.

3.2.3 Ajustement

- **statistiques d'ajustements :**

L'ajustement du modèle a été réalisé à l'aide des fonctions `glmer` de R, et `GLIMMIX` de SAS dont les algorithmes ont tous deux convergé.

Les statistiques de test obtenues avec R et SAS sont :

glmer		GLIMMIX	
AIC	130,25	pseudo-AIC	135,28
BIC	167,68	pseudo-BIC	163,31
déviante	90,25	-2log(pseudo-L)	95,28

Statistiques de test obtenues avec R et SAS.

On remarque que les résultats obtenus avec les deux fonctions sont assez voisins.

- **estimations des paramètres des effets fixes :**

Afin de pouvoir comparer aisément les résultats renvoyés par les deux logiciels, nous avons eu recours à la fonction `relevel` de R. Elle permet de réordonner les niveaux d'un facteur, et ainsi de choisir les effets estimés à 0.

Les estimations obtenues par `GLIMMIX` et `glmer` (cf *tableaux 3 et 4*) ne sont pas identiques, à cause des algorithmes de calcul propres à chaque fonction, mais elles sont dans l'ensemble très proches.

- **structure de covariance :**

D'après notre modèle, la matrice de variance covariance de y est définie par la matrice bloc-diagonale suivante (le nombre de bloc est égal au nombre de parcelle, et la taille d'un bloc est de $j \times j$ où j est le nombre d'horizon):

On a donc : $\widehat{\alpha(\phi)} \gg 1$.

Cela confirme donc bien le fait que l'on soit dans un cas de surdispersion des données.

Remarque :

Le résultat de la statistique du test de surdispersion (énoncé dans la partie théorique) est de: $|T| = 0.697$. Donc $|T| < 1.96$ où 1.96 est le quantile de la loi $\mathcal{N}(0,1)$ pour $\alpha = 0.05$. D'après ce test on devrait accepter \mathcal{H}_0 , c'est-à-dire considérer que les données n'étaient pas surdispersées.

Toutefois, il se peut que ce test soit particulièrement peu puissant, notamment car la taille de l'échantillon étudié est seulement de $n = 48$.

L'estimation du paramètre de dispersion étant nettement supérieure à 1, on conclut tout de même que l'on est dans un cas de surdispersion des données.

- **problème rencontré lors de l'estimation de l'effet aléatoire :**

Pour l'étude des densités de vers blancs dans les cultures de riz au mois de Mars, l'estimation de l'effet aléatoire bloc : T, n'a pas posé de difficulté.

En revanche, pour certaines variables, il est arrivé que la procédure GLIMMIX de SAS renvoie le message suivant :

*« L'estimation de la matrice G n'est pas définie positive.
Note : la matrice de covariance est la matrice nulle. »*

Ceci est dû au fait que l'estimation de l'effet aléatoire obtenue était si petite que la procédure la considère comme nulle. En effet, par exemple pour l'indice de richesse dans les cultures de maïs au mois de mars, l'estimation de l'effet aléatoire obtenue était:

Covariance Parameter Estimates			
Param de cov	Sujet	Valeur estimée	Erreur type
bloc*T	parcelle	6.24E-19	.

Estimation des paramètres de la matrice de variance-covariance avec GLIMMIX.

La fonction ne calcule pas l'erreur car il considère l'estimation de l'effet aléatoire trop faible.

Dans un tel cas, la procédure GLIMMIX poursuit ses calculs en considérant qu'il n'a pas d'effets aléatoires dans le modèle. (En effet, les estimations de moyennes renvoyées sont égales aux estimations du même modèle sans les effets aléatoires).

(On remarque par ailleurs ici que la variance résiduelle n'est pas donnée car on ne se trouve pas ici dans un cas de surdispersion).

3.2.4 Validation

L'analyse des résidus a permis de valider notre modèle. Les conclusions des tests obtenues avec les deux logiciels étant les mêmes, seuls les résultats de la procédure GLIMMIX figurent en annexe (cf *tableau 5*, et *graphe 3*).

Pour que le modèle soit correctement ajusté, les résidus doivent suivre une loi normale, être d'espérance nulle, et de variances homogènes, et être indépendants entre eux.

Ces quatre hypothèses ont été vérifiées, et ce, au moyen de différents tests :

- normalité des résidus :

Elle a été vérifiée par les tests de Shapiro-Wilk, Kolmogorov-Smirnov, et Anderson-Darling.. Graphiquement, cela peut s'observer aussi en traçant un quantile-quantile plot.

- tendance centrale des résidus :

Cette hypothèse a été confirmée au moyen du test de student.

- Homogénéité des variances et indépendances des résidus:

On a observé pour cela le graphe des résidus de déviance en fonction des valeurs prédites. Comme l'amplitude était relativement constante, et les valeurs centrées en zéro, on en a déduit que les variances étaient assez homogènes. On ne dénote aucune organisation particulière dans la disposition des résidus, on peut donc considérer qu'ils étaient indépendants.

Le test de Levène a également permis de prouver l'homogénéité des variances.

3.2.5 Analyse des effets fixes

Comme énoncé antérieurement, pour pallier l'impossibilité de réaliser une ANOVA avec notre modèle dans R, nous avons procédé à une succession de test de rapport de vraisemblance sur des modèles emboîtés. Détaillons ici la première étape :

Soient \mathcal{M}_1 et \mathcal{M}_2 les modèles GLMM (avec distribution quasi-poissonnienne) définis tels que :

$$\mathcal{M}_1 : \log(\mu_{ijk}) = \nu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + B_{ik}$$

$$\mathcal{M}_2 : \log(\mu_{ijk}) = \nu + \alpha_i + \beta_j + \gamma_k + B_{ik}$$

où $\forall i, j, k$ avec $i = 1, \dots, 8$; $j = 1, 2$ et $k = 1, \dots, 3$:

Alors, on déduit du test du rapport de vraisemblance que réalise $anova(\mathcal{M}_1, \mathcal{M}_2)$ que l'effet de l'interaction $(\alpha\beta)_{ij}$ (ie : T : H) était significative car la p-value indiquée est de : 0.009379.

On choisit donc de garder cette interaction dans notre modèle, et on procède pareillement avec les effets fixes. Par exemple, pour tester l'effet du bloc, on effectue un test du rapport des vraisemblances entre les deux modèles qui suivent:

$$\mathcal{M}_1 : \log(\mu_{ijk}) = \nu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + B_{ik}$$

$$\mathcal{M}_3 : \log(\mu_{ijk}) = \nu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + B_{ik}$$

On obtient cette fois-ci une p-value de 0,206 , ce qui permet de déduire que l'effet du bloc n'était pas significatif.

En réitérant cette méthode pour les autres effets fixes, on est parvenu aux conclusions suivantes : les effets du **traitement**, de l'**horizon**, et de l'interaction **traitement : horizon** étaient significatifs.

Nous avons également retrouvé ces résultats en utilisant une autre méthode : la sélection de modèle par le critère AIC (Burnham et Anderson 1998) :

$$AIC = -2 \log(L) + 2k$$

où L est la vraisemblance du modèle considéré, et k le nombre de paramètre

Ici, notre modèle se base sur une quasi-vraisemblance, et l'AIC correspondant s'appelle alors le Quasi-AIC (Anderson et Burnham 1994) et est défini par :

$$QAIC = -\frac{1}{a(\phi)} \times 2 \log(QL) + 2k$$

où QL est la quasi-vraisemblance de notre modèle, et $a(\phi)$ le paramètre de dispersion.

Nous avons suivi « manuellement » la même méthode itérative que celle de la fonction step de R (qui ne prend pas malheureusement pas en compte les modèles de type GLMM) afin de sélectionner le modèle meilleur modèle, c'est-à-dire celui minimisant l'AIC, en enlevant au fur et à mesure les effets fixes de notre modèle.

Le modèle sélectionné par cette méthode est le suivant :

$$\mathcal{M} : \log(\mu_{ijk}) = \nu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + B_{ik}$$

On est parvenu ainsi aux mêmes conclusions sur les effets fixes significatifs qu'avec les tests de rapport de vraisemblance.

Les résultats obtenus avec GLIMMIX par sélection de modèle via l'AIC sont identiques. On pouvait s'attendre à cela étant donné la similitude observée entre les statistiques d'ajustement estimées avec glmer et GLIMMIX.

Avec sas, l'analyse des effets fixes s'est révélée beaucoup plus simple et rapide. La procédure GLIMMIX permet de réaliser une ANOVA de type III, qui est la méthode la plus adaptée à notre problème. Le tableau renvoyé est le suivant :

Tests de type III des effets fixes				
Effet	DDL Num.	DDL Res.	Valeur F	Pr > F
bloc	2	14	0.89	0.4329
T	7	14	0.29	0.9475
H	1	16	29.74	<.0001
T*H	7	16	1.17	0.3703

Table d'ANOVA obtenue avec GLIMMIX

Avec cette méthode, le seul effet qui apparaît comme significatif, au seuil $\alpha = 0.005$, est celui de l'**horizon**.

On observe donc que les résultats obtenus en utilisant des tests de rapport de vraisemblance (ou AIC), sont totalement différents de ceux obtenus par l'analyse de variance avec SAS.

Les résultats renvoyés par l'ANOVA de SAS semblent tout de même plus pertinents :

- En effet, l'ANOVA est la méthode la plus adaptée pour mettre en évidence la significativité des effets fixes.
- De plus, il est peut-être possible que les résultats obtenus en utilisant la vraisemblance du modèle soient faussés à cause de l'estimation de celle-ci.
- Enfin, Pinheiro et Bates (2000) déconseillent la méthode du rapport de vraisemblance pour tenter de mettre en évidence les effets fixes significatifs dans un modèle mixte.

En adoptant donc les résultats obtenus avec l'analyse de variance, on peut conclure ici que : seul l'effet de l'**horizon** était significatif des densités de vers blancs.

3.2.6 Estimation des moyennes

Les estimations des moyennes ont été obtenues grâce à la formule suivante :

$$\widehat{\mu}_{ijk} = \exp(\hat{\nu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + (\widehat{\alpha\beta})_{ij} + \widehat{B}_{ik})$$

$\forall i, j, k$ avec $i = 1, \dots, 8$; $j = 1, 2$ et $k = 1, \dots, 3$.

Dans R, la fonction *fitted* permet d'obtenir ces estimations directement. Tandis qu'avec SAS, les estimations des moyennes par moindres carrés (LS Means) fournit par GLIMMIX correspondent en réalité aux : $\log(\widehat{\mu}_{ijk})$. Il faut donc appliquer l'inverse de la fonction de lien (ici *exp*) à ces résultats pour retrouver les estimations souhaitées.

3.2.7 Comparaisons de moyennes

L'ANOVA met en évidence l'existence (ou non) d'une différence entre les moyennes au sein d'un même groupe. S'il existe bien une différence, on utilise alors des comparaisons multiples de moyennes afin de déterminer précisément quelles sont les moyennes qui diffèrent significativement les unes des autres.

Toute comparaison de moyennes s'écrit sous la forme d'un contraste, (Letourmy 1999) c'est-à-dire une combinaison linéaire des moyennes des traitements dont la somme des coefficients est nulle :

$$C = \sum_i a_i \mu_i \text{ est un contraste si : } \sum_i a_i = 0$$

Ils existent un grand nombre de méthodes de comparaisons multiples qui diffèrent soit par les contrastes étudiés, soit par le risque d'erreur qu'elles contrôlent : Bonferroni, Newman-Keuls, Dunnett, Duncan, méthode des contrastes...

Nous avons choisi ici d'avoir recours à la méthode des contrastes qui permet de contrôler le risque situé au niveau de la comparaison.

Cette méthode effectue le test suivant :

$$H_0 : C = 0$$
$$H_1 : C \neq 0$$

Pour effectuer ce test, on peut utiliser un test de Student comme le fait la procédure GLIMMIX, ou un test de Fisher.

Dans cette étude, nous nous sommes intéressés aux contrastes élémentaires, c'est à dire les contrastes du type : $C = \mu_i - \mu_j$ (comparaison de moyennes deux à deux).

Notre projet consistant à analyser l'effet des différents traitements sur les macrofaunes du sol, les comparaisons de moyennes n'ont été effectuées que lorsque l'ANOVA avait dans un premier temps démontré que des différences significatives existaient entre les traitements. Pour les densités de vers blancs en mars dans le dispositif riz pluvial, l'effet du traitement n'était pas significatif. Nous allons donc ici traiter une autre variable pour illustrer les comparaisons de moyennes deux à deux: le poids des larves *Apicencya Waterlotii* (essai n°1 en laboratoire). Pour cette variable, les effets du traitement mais aussi de la date étaient significatifs. Nous allons donc examiner les comparaisons des moyennes calculées par traitement et par date.

Comme spécifié dans la partie matériel, cette étude n'a pu être réalisée avec R. En revanche, cela s'effectue aisément avec GLIMMIX : en précisant `pdiff` dans la fonction `lsmeans`.

Une partie du tableau que l'on obtient en sortie se trouve en annexe (cf *tableau 6*), avec l'interprétation des différences significatives entre les traitements.

3.3 Présentation synthétique des résultats et interprétations

Dans cette section nous allons présenter brièvement les résultats des analyses de variance effectuées sur chaque variable, et commenter lorsque cela est intéressant les moyennes estimées (ces moyennes estimées figurent en annexe). Dans tous les tableaux qui suivent, les effets significatifs de l'ANOVA sont représentés de la manière suivante :

significatif = effet significatif au seuil $\alpha=5\%$ significatif = effet significatif au seuil $\alpha=10\%$

3.3.1 Densités de vers blancs

Pour chacun des trois dispositifs, on a observé les faits suivants :

- Les densités de vers blancs étant trop faibles dans la litière et l'horizon 0-30 cm, notre étude se base seulement sur les prélèvements des horizons de 0 à 20 cm de profondeur,
- Les données étaient surdispersées,
- Les densités étaient nettement plus élevées au mois de mars qu'au mois d'août. C'est pourquoi nous avons examiné si la date pouvait avoir un impact sur le nombre de vers blancs prélevé, pour chacun des deux horizons au moyen du même modèle en remplaçant seulement le facteur horizon par le facteur date.

	effet étudié	traitement	horizon	bloc	traitement:horizon
Riz	mars	-	significatif	-	-
	août	significatif	-	-	-
Maïs	mars	significatif	significatif	significatif	-
	août	-	significatif	significatif	-
Plantes seules	mars	-	significatif	-	-
	août	-	-	-	-

Résultats des analyses de variance sur les densités de vers blancs

Résultat des comparaisons de moyennes effectuées lorsque le traitement était significatif :

Pour les données concernant le dispositif riz en mars, on a déduit qu'il y avait des différences significatives entre les traitements suivants : R2 et R3, R3 et R4, R3 et R5, R3 et RLH, et entre R3 et RS. Il semblerait donc que le traitement R3 (**Riz+Cléome+Tagète+Cosmos**) ait présenté des différences significatives avec la plupart des autres traitements. Tandis que pour les densités des cultures de maïs au mois de mars, il semblerait que se soient les traitements M2 (**Maïs+Eleusine+Crotalaire**) et M3 (**Maïs+Cléome+Tagète+Cosmos**) qui aient présenté des différences significatives avec la plupart des autres traitements.

Dans le dispositif riz pluvial, l'effet de la date est apparu significatif seulement pour l'horizon 0-10cm.

Par ailleurs, dans les cultures de maïs, la date était significative pour chacun de ces deux horizons. Cela signifie qu'il existait bien une relation entre le nombre de vers blancs et le mois au cours duquel ils ont été prélevés.

Enfin, pour les données du dispositif plantes seules : l'effet de la date était significatif (ainsi que l'interaction traitement : date au seuil de 10%) seulement pour l'horizon 0-10 cm.

Cela signifie que pour les trois dispositifs, il existait bien une relation entre le nombre de vers blancs et le mois au cours duquel ils ont été prélevés mais cela était significatif principalement pour les données de l'horizon 0-10cm.

3.3.2 Diversité de la macrofaune du sol et des Pitfall trap

Les résultats des ANOVA effectuées sur les indices de diversité de la macrofaune du sol figurent dans le *tableau 7* en annexe. L'étude globale de ce tableau permet de déduire que l'effet de l'horizon était significatif quelque soit le dispositif, la date de prélèvement, et l'indice de diversité considérés.

Par ailleurs, l'effet du bloc apparait également significatif des indices de Shannon et Gini-Simpson dans les cultures de maïs au mois d'août, ainsi que pour l'indice ce Gini-Simpson dans le dispositif plantes seules au mois de mars.

Enfin, dans le dispositif riz pluvial au mois d'août, le traitement était significatif de l'indice de Shannon (au seuil $\alpha = 10\%$). Les comparaisons de moyennes ont alors mis en évidence le fait que des différences significatives existaient principalement entre l'indice de Shannon des traitements R3 (**Riz + Cléome + Tagète + Cosmos**) et R2 (**Riz + Eleusine + Crotalaire**), et celui des autres traitements.

Pour les indices de diversité dans les Pittfall trap, les résultats des analyses de variances figurent dans le tableau ci-dessous :

	effet étudié	traitement	bloc
Riz	richesse	-	-
	shannon	-	significatif
	Gini-Simpson	-	significatif
Maïs	richesse	-	-
	shannon	-	-
	Gini-Simpson	-	-
Plantes seules	richesse	-	-
	shannon	significatif	-
	Gini-Simpson	significatif	-

Résultats des analyses de variance sur les indices de diversité, dans les Pitfall trap.

Pour l'indice de Shannon et celui de Gini-Simpson, les différences significatives sont apparues entre les traitements :

- Cléome viscosa et brachiaria mulato,
- Cléome viscosa et radis fourrager,
- Radis fourrager et vesce,

- *Brachiaria mulato* et vesce.

Les corrélations entre les densités de vers blancs et celles des autres insectes étaient très faibles, et il semblerait qu'il n'y ait aucun insecte en particulier dont la densité soit liée à celle des vers blancs.

Nous avons également regardé les corrélations existantes entre les différents indices de diversité et les densités de vers blancs. Le coefficient de corrélation le plus élevé avec le nombre de vers blancs est celui obtenu avec l'indice de richesse (entre 0.35 et 0.65 en fonction du dispositif). Toutefois, ces résultats sont trop faibles pour en déduire une réelle dépendance entre les deux variables. Les indices de Shannon et Gini-Simpson sont, quand à eux, fortement corrélés.

3.3.3 Production de riz et dégâts observés

D'après l'analyse de variance des données de production, l'effet bloc seulement était significatif. De plus, concernant le taux de poquet attaqués sur les poquets restants : seuls les effets de la date et du bloc étaient significatifs. Ainsi, dans ces deux cas, les effets des traitements n'ont pas été trouvés significatifs ce qui pourrait paraître étrange étant donné les différences assez marquées que l'on peut voir dans les estimations de moyennes par traitement pour la production (cf *tableau ci-dessous*) ou le taux de poquets attaqués (cf *tableau 8* en annexe).

production moyenne par traitement	
R5	18,6
R3	17,8
RS	16,7
R1	16,4
R4	16,1
R2	9,2
RSDH	9,1
RLH	5,6

Classement des traitements en fonctions de la production de riz moyenne estimée (en kg)

En réalité, dans le cas particulier des données de production, l'estimation de la variance résiduelle donnée par la table d'ANOVA est : $\sigma_e^2 = 57.414$. Ainsi :

$$Var(\bar{y}_1 - \bar{y}_2) = \sigma_e^2 \left(\frac{1}{n} + \frac{1}{n} \right) = \frac{2 \sigma_e^2}{n}$$

où \bar{y}_1 et \bar{y}_2 sont deux estimations de moyennes du tableau ci-dessus, et le nombre de répétitions $n = 3$.

Par conséquent l'intervalle de confiance de $\bar{y}_1 - \bar{y}_2$ est très large :

$$IC_{\bar{y}_1 - \bar{y}_2} = \overline{\bar{y}_1 - \bar{y}_2} \pm 1.96 \sqrt{\frac{2 \sigma_e^2}{n}} \approx \overline{\bar{y}_1 - \bar{y}_2} \pm 12.126$$

Par exemple pour $\bar{y}_1 = \overline{\bar{y}_{R3}}$ et $\bar{y}_2 = \overline{\bar{y}_{RS}}$, alors :

$$IC_{\overline{y_{R3}} - \overline{y_{RS}}} \approx 17.8 - 16.7 \pm 1.96 \sqrt{\frac{2 \sigma_e^2}{n}} \approx [-11.026 ; 13.226]$$

Ainsi, on peut conclure qu'il y a eu un important problème de variabilité dans les mesures tant pour la production que pour le taux de poquets attaqués.

Par ailleurs, après l'analyse des résidus de la variable « taux de poquets attaqués », nous a permis de constater que le modèle ne s'ajustait pas correctement aux données. Nous avons donc procédé à une transformation des données à l'aide de la fonction *logit*. Pour analyser ces données transformées, nous avons modifié le modèle en choisissant une distribution gaussienne, et un lien identité

3.3.4 Effets biocide des associations de plantes sur les larves

Voici le bilan des résultats de l'ANOVA :

	effet étudié	traitement	date	traitement:date
Heteronychus Arator (a)	essai 1	-	significatif	-
	essai 2	-	significatif	-
Apicencya Waterlotii (b)	essai 1	significatif	significatif	significatif
	essai 2	-	significatif	significatif
Heteroconus paradoxus (c)	essai 1	significatif	significatif	-
	essai 2	-	significatif	-
	essai 3	-	significatif	-

Résultats des analyses de variance des poids de larves.

On remarque que quels que soient les larves et les essais considérés, l'effet de la date était significatif. Cela signifie donc que le poids des larves a varié de façon significative suivant les différentes dates de prélèvement. Pourtant, cela ne semble pas toujours causé par le traitement. En effet, le traitement était seulement significatif du poids des larves (b) *Apicencya Waterlotii* et (c) *Heteroconus paradoxus* des essais 1.

Pour chacun des essais sur les larves (b), on remarque tout de même que l'interaction traitement : date était significative. On en conclut donc que le poids de ces larves a varié significativement au cours du temps, et que cela est lié aux traitements employés dans l'expérience. Les comparaisons de moyenne mettent en évidence des différences significatives principalement entre le poids des (b) élevées avec le traitement **Cajanus**, et le poids de ces mêmes larves dans les milieux correspondants aux traitements **témoin**, **Eleusine**, **Desmodium**, et **Radis fourrager**.

Pour les larves (c) de l'essai 1, il semblerait plutôt que les différences significatives aient été observées entre le traitement **témoin**, et les traitements **Eleusine**, **Cajanus**, et **Radis fourrager**.

3.4 Limites de ce travail, et suggestions d'améliorations

Abordons dans un premier temps, les questions se rapportant au dispositif expérimental même.

L'étude globale de l'effet des différents traitements employés s'est malheureusement révélée assez décevante. D'après la forte variabilité des mesures de production de riz, on peut supposer que le développement du riz a peut-être été altéré par la précédente culture (l'avoine), ou par les différentes couvertures végétales. Il pourrait être intéressant de pouvoir examiner les données de production de maïs, afin d'observer si la variabilité est aussi importante pour cette variable.

De plus, au sein même d'une population de vers blancs, on détecte bien entendu des ravageurs, mais également une grande partie de non ravageurs et notamment des ingénieurs du sol (ie : des larves bénéfiques au biofonctionnement du sol). Cependant, les données de vers blancs dont nous disposons ici ne renseignaient pas sur le type de vers blancs relevés. Il serait donc important par la suite de pouvoir appliquer les méthodes qui ont été utilisées ici en traitant seulement les densités de vers blancs ravageurs.

Par ailleurs, d'un point de vue statistique, on pourrait suggérer plusieurs améliorations.

En effet, il pourrait être intéressant tout d'abord de comparer le modèle dont la distribution est quasi-poissonnienne à un modèle avec une autre distribution permettant également de tenir compte de la surdispersion des données telle que la distribution binomiale négative. En effet, cette distribution a été utilisée par Dean et Lawless (1989) afin de répondre aux problèmes de surdispersion liées aux modèles de Poisson, et possède les propriétés suivantes :

$$E(y_i) = \mu_i \quad \text{et} \quad Var(y_i) = \mu_i + \frac{1}{k} \mu_i^2 \quad \text{où } \frac{1}{k} \text{ est le paramètre de dispersion.}$$

Cependant, la comparaison de ces deux modèles par des statistiques d'ajustements telles que la déviance, ou le critère AIC est sans doute complexe étant donné que les méthodes d'estimation des vraisemblances sont différentes : maximum de quasi-vraisemblance pour la première, et maximum de vraisemblance pour la seconde.

En outre, comme les horizons 20-30 cm et la litière contenaient une grande majorité de valeurs nulles, nous avons choisi ici de ne considérer que les densités des horizons 0-10 cm et 10-20 cm, mais il pourrait être intéressant de prendre en compte tous les horizons en utilisant une distribution plus adaptée. En effet, le modèle ZIP (zero-inflated poisson) par exemple, permet de traiter les données possédant un excès de zéros (Allain et Brenac 2001). Il relève d'un mélange d'une loi de Poisson et d'une loi de Dirac en 0.

Il existe également un modèle binomial négatif modifié en zéro (ZINB) (Shankar, Milton et Mannering 1997). Il serait pertinent de voir s'il est possible de modéliser un modèle mixte avec une telle distribution.

Enfin, concernant les poids de larves mesurés pour mettre en évidence l'effet biocide des plantes de couverture, nous avons choisi d'affecter un poids nul à une larve morte, et ainsi nous avons pu réutiliser le même type de modèle que pour les variables précédentes. Cependant, il serait peut-être plus judicieux ici de traiter ces données en tant que données de survie censurées, au moyen par exemple d'un modèle de Cox.

4. Conclusion

Bilan des résultats et limites de ce travail

Cette étude a permis de comparer les effets de diverses plantes associées à une culture principale telle que le riz ou le maïs sur le contrôle des vers blancs. Pour la réaliser, les modèles linéaires généralisés mixtes se sont révélés particulièrement adaptés au dispositif expérimental qui avait été mis en place préalablement à Madagascar. De plus, la prise en compte de la surdispersion des données, par l'intermédiaire d'une distribution quasi-poissonnienne, s'est avérée nécessaire. Toutefois, d'autres distributions auraient pu être envisagées telles que la loi binomiale négative, adaptée elle aussi à la modélisation de la surdispersion, ou bien encore le modèle ZIP qui permettrait de pallier l'excès de densités nulles qu'il est possible de rencontrer dans certains horizons.

La première étape de notre projet a été de comparer les effets de différentes plantes de service sur les densités de vers blancs, ainsi que sur la biodiversité de la macrofaune du sol. Il faut reconnaître que les effets des traitements se sont révélés le plus souvent assez décevants. Il semblerait toutefois qu'au cours de certaines occasions le traitement n'ait pas été négligeable. Dans ces derniers cas, les densités de vers blancs présents dans les associations de plantes « Cléome + Tagète + Cosmos » et « Eleusine + Crotalaire » présentaient alors des différences significatives avec les effectifs de vers blancs élevés dans les autres milieux. Par ailleurs, il pourrait être approprié d'examiner uniquement les effectifs des vers blancs ravageurs.

Dans une seconde étape, nous avons étudié le rendement de riz, ainsi que les dégâts observés sur cette culture principale. Toutefois, là encore, les traitements ne se sont pas révélés significatifs, et aucune corrélation n'a pu être établie entre ces résultats, et les densités de vers blancs ou les indices de diversité. Nous avons également remarqué un important problème de variabilité dans les mesures, notamment pour la production. Ceci pourrait être causé par une altération du développement des cultures de riz.

Enfin, dans une troisième et dernière étape, nous avons tenté de mettre en évidence l'effet biocide des associations de plantes de couverture sur plusieurs variétés de vers blancs. Dans certaines situations particulières, lorsque les poids des larves étaient liés aux traitements, sont apparues des différences significatives entre le milieu « témoin » et le milieu contenant des résidus de « Cajanus ». Il deviendrait alors judicieux, par la suite, d'explorer les possibilités d'un modèle peut-être plus adapté que celui qui a été choisi ici, en considérant les données de poids de larves comme données de survie censurées.

Par ailleurs, d'un point de vue pragmatique, la procédure GLIMMIX du logiciel SAS s'est avérée tout au long de cette étude, particulièrement adéquate au problème posé, surtout pour le test des contrastes, contrairement à la fonction glmer de R qui s'est révélée source de plusieurs difficultés.

Expérience personnelle

A première vue, l'objectif même du projet de l'ATP Omega 3, au sein duquel s'inscrit ce stage, à la croisée de l'écologie et de l'économie vivrière des populations de Madagascar, m'a paru excessivement pertinent. Il a été très valorisant pour moi que le travail fourni ait une application concrète *a posteriori*. De plus, grâce à cet objectif motivant, une progression personnelle, au cours de cette recherche, est toujours possible malgré les différents problèmes rencontrés de nature agronomique, statistique ou encore logicielle.

Par contre, un des points qui m'a plutôt ennuyé est de n'avoir trouvé que de rares références se rapportant à l'application des modèles linéaires généralisés mixtes utilisés avec le logiciel R. Toutefois, à cette occasion, j'ai pu ainsi me perfectionner dans l'utilisation du logiciel SAS.

Bien que légèrement désappointée par les résultats peu significatifs de l'expérience réalisée, tant sur le terrain qu'au cours de ce présent travail, ce fut également pour moi intéressant et instructif de construire toute une méthodologie pour parvenir à étudier ce dispositif agronomique spécifique.

Enfin, le travail en équipe m'a beaucoup apporté. Tout d'abord, les discussions avec les autres biostatisticiens m'ont permis de comprendre plusieurs points théoriques essentiels. De plus, les nombreux échanges avec les agronomes, qui eux connaissaient le terrain et le dispositif expérimental étudié, se sont révélés tant utiles dans la rédaction de ce rapport, que conviviaux.

Perspectives d'ouverture

Ces travaux soulèvent finalement de nouvelles pistes. Le principal est d'essayer désormais d'élaborer un modèle dynamique qui puisse prédire, en fonction des plantes de couvertures utilisées, les effectifs ou les poids de vers blancs. On peut suggérer que cela pourrait s'effectuer à l'aide d'un modèle déterministe par équations différentielles. Un tel modèle permettrait alors d'élaborer un dispositif expérimental argumenté pour les recherches de la saison suivante.

Bibliographie

- Allain E. et Brenac T. (2001) Modèles linéaires généralisés appliqués à l'étude des nombres d'accidents sur des sites routiers : le modèle de Poisson et ses extensions, *Recherche Transports Sécurité* 72, 3-18. Elsevier Ltd.
- Anderson D.R., et Burnham K.P., (1994) AIC model selection in overdispersed capture-recapture data, *Ecology*, 75(6), pp. 1780-1793 Ecological Society of America.
- Breslow, N. E. & Clayton, D. G. (1993) Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 88, 9–25.
- Burnham, K. P. et Anderson, D. R. (1998) Model selection and inference. A practical information-theoretic approach. Springer-Verlag, New York.
- Dean, C. B. and Lawless, J. F. (1989) Tests for detecting overdispersion in Poisson regression models, *Journal of the American Statistical Association* 84, 467-472.
- Dean C. B. (1992) Testing for overdispersion in Poisson and binomial regression models. *Journal of the American Statistical Association*, 87, 451-457.
- Dobson, A. (1990) An Introduction To Generalized Linear Models, London: Chapman and Hall, 60.
- Engel, B. and Keen, A. (1992) A Simple Approach for the Analysis of Generalized Linear Mixed Models, LWA-92-6, *Agricultural Mathematics Group (GLW-DLO)*. Wageningen, The Netherlands.
- Gilmour, A.R., Anderson, R.D., and Rae, A.L. (1985) The Analysis of Binomial Data by Generalized Linear Mixed Model, *Biometrika*, 72, 593-599.
- Green, P. J. (1987) Penalized Likelihood for General Semi-Parametric Regression Models, *International Statistical Review*, 55, 245-259.
- Henderson, C.R., O. Kempthorne, C. Searle, and C. VonKrosig. (1959). The estimation of environmental and genetic trends from records subject to culling. *Biometrics* 15, 192-218.
- Jayaraman K. (1999) Manuel de statistique pour la recherche forestière, *Kerala Forest Research Institute Peechi, Thrissur, Kerala (Inde)*, p 110.
- Lavergne C., Trottier C. (2000) Sur l'estimation dans les modèles linéaires généralisés mixtes, *Revue de Statistiques Appliquées*, XLVIII (1).
- Letourmy P., revision Gozé E. (1999) Expérimentation agronomique planifiée (support de cours). Cirad, Montpellier.
- Littell, Ramon C., George A. Milliken, Walter W. Stroup, Russell D. Wolfinger, and Oliver Schabenberger. (2006) SAS for Mixed Models, Second Edition. Cary, NC: SAS Institute Inc.
- Marque S. (2003) Prise en compte de la surdispersion par des modèles à mélange de Poisson, Thèse de doctorat de l'Université de Bordeaux 2

- McCullagh, P. et J. A. Nelder. (1989) *Generalized linear models*, second edition. Chapman and Hall.
- McCulloch, Charles E. et Shayle R. Searle. (2001) *Generalized, linear and mixed models*. New York NY : John Wiley and Sons.
- Nelder, J. A. et R. W. M. Wedderburn. (1972) *Generalized linear models*, *Journal of the Royal Statistical Society, Series A*, vol. 135, p. 370-384.
- Pavoine, S. (2005) *Méthodes Statistiques pour la Mesure de la Biodiversité*, Thèse de doctorat de l'Université Claude Bernard Lyon I.
- Pinheiro, J. C. & D. M Bates. (2000) *Mixed-Effects Models in S and S-PLUS*. New York: Springer.
- Rakotoniadranaivo M. (2009) *Evolution de la macrofaune tellurique suivant divers systèmes de culture et effets particuliers sur les vers blancs*.
- Ratnadass A., Fernandes P., Avelino J., Letourmy P., Deberdt P., Naudin K., Rhino B., Deguine J.P., Babin R., Gozé E., Tixier P., Laurent J.B. (2010) *Rapport Scientifique 2009 de l'ATP dirigée Omega 3*.
- Schall, R. (1991) *Estimation in Generalized Linear Models with Random Effects*, *Biometrika*, 78, 719-727.
- Searle, S. R., Casella, G., and McCulloch, C. E. (1992) *Variance Components*. New York: Wiley, p456-475
- Shankar, V., J. Milton et F. Mannering (1997) *Modeling accident frequencies as zeroaltered probability processes : an empirical inquiry*, *Accident Analysis and Prevention*, 29(6), 829-837, 1997.
- Trottier C. (1998) *Estimation dans les modèles linéaires généralisés à effets aléatoires*, Thèse de doctorat de l'Institut National Polytechnique de Grenoble
- Wolfinger, R., and M. O'Connell. (1993), *Generalized Linear Models : A Pseudo-likelihood Approach*, *Journal of Statistical Computation and Simulation* 48:233-243.

Annexes

Propriétés des trois distributions utilisées

Tableaux

Tableau 1 : table d'analyse de la variance pour un dispositif split-plot

Tableau 2 : analyse exploratoire avec proc univariate

Tableau 3 : estimation des effets fixes avec GLIMMIX

Tableau 4 : estimation des effets fixes avec *glmer*

Tableau 5 : analyse des résidus avec GLIMMIX (excepté le test de Levène obtenu avec R)

Tableau 6 : Table des comparaisons de moyenne deux à deux (obtenue avec GLIMMIX).

Tableau 7 : Résultats de l'ANOVA des indices de diversité de la macrofaune du sol

Tableau 8 : Classement des traitements en fonctions de la moyenne de l'estimation du taux de poquets attaqués sur le nombre de poquets restants

Graphiques

Graphe 1 : histogramme des densités de vers blancs

Graphes 2 : boxplot des densités de vers blancs dans les cultures de riz en mars

Graphe 3 : graphes des résidus (pour la variable densité de vers blancs dans le dispositif riz (mars))

Estimations des densités de vers blancs

Estimations des indices de diversité moyens de la macrofaune du sol

Estimations des indices de diversité moyens de la macrofaune des Pitfall trap par traitement

Propriétés des trois distributions utilisées

- Distribution gaussienne

Lien : identité

$$\text{Densité de probabilité : } f(y) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right) \exp \left[-\frac{1}{2\sigma^2} (y - \mu)^2 \right]$$

$$\text{Log-vraisemblance : } l(\mu, \sigma^2; y) = \left(-\frac{1}{2\sigma^2} \right) (y - \mu)^2 - \frac{1}{2} \log(2\pi\sigma^2)$$

- Distribution poissonnienne

Lien : log

$$\text{Densité de probabilité : } f(y) = \frac{\lambda^y e^{-\lambda}}{y!}$$

$$\text{Log-vraisemblance : } l(\lambda; y) = y \log(\lambda) - \lambda - \log(y!)$$

- Distribution binomiale

Lien : *logit* définie telle que : $\text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right)$

$$\text{Densité de probabilité : } f(y) = \binom{n}{y} \pi^y (1 - \pi)^{n-y}$$

$$\text{Log-vraisemblance : } l(\pi; y) = y \log\left(\frac{\pi}{1-\pi}\right) + n \log(1 - \pi) + \log\binom{n}{y}$$

	NORMAL	POISSON	BINOMIAL
$\mathbb{E}(Y)$	μ	λ	$n\pi$
$\theta(\mu)$	μ	$\log(\lambda)$	$\text{logit}(\pi)$
$a(\phi)$	σ^2	1	1
$V(\mu)$	1	λ	$n\pi(1-\pi)$
$\text{Var}(Y)$	σ^2	λ	$n\pi(1-\pi)$
R_μ	I	$\text{diag}(\lambda_i)$	$\text{diag}(n\pi_i(1-\pi_i))$
A	$\sigma^2 I$	I	I

(Littell, et al. 2006 p 532-533).

Tableaux

Tableau 1 : table d'analyse de la variance pour un dispositif split-plot

Variation	SCE	dl	CM	F
Facteur A	$\sum_i nJ (Y_{i..} - Y_{...})^2$	I - 1	SCA/(I-1)	CMA/CMR1
Bloc	$\sum_k IJ (Y_{..k} - Y_{...})^2$	n - 1	SCb/(n-1)	CMB/CMR1
Résiduelle 1	$\sum_{ik} J (Y_{i.k} - Y_{i..} - Y_{..k} + Y_{...})^2$	(I-1)(n-1)	SCR1/ (I-1)(n-1)	
Facteur B	$\sum_j nI (Y_{.j.} - Y_{...})^2$	J - 1	SCB/ (J-1)	CMB/CMR2
Interaction A x B	$\sum_{ij} n (Y_{ij.} - Y_{i..} - Y_{.j.} + Y_{...})^2$	(I-1) (J-1)	SCI/(I-1)(J-1)	CMI/CMR2
Résiduelle 2	$\sum_{ijk} (Y_{ijk} - Y_{i.k} - Y_{.j.} + Y_{...})^2$	I (J-1)(n-1)	SCR2/I(J-1)(n-1)	

(Letourmy 1999)

Tableau 2 : analyse exploratoire avec proc univariate

Moments			
N	96	Somme des poids	96
Moyenne	3.3125	Somme des observations	318
Ecart-type	4.80857239	Variance	23.1223684
Skewness	2.26788137	Kurtosis	5.89320901
Somme des carrés non corrigée	3250	Somme des carrés corrigée	2196.625
Coeff Variation	145.164449	Std Error Mean	0.49077286

Mesures statistiques de base			
Tendance centrale		Variabilité	
Moyenne	3.312500	Ecart-type	4.80857
Médiane	1.000000	Variance	23.12237
Mode	0.000000	Intervalle	25.00000
		Ecart interquartile	4.50000

Tests de tendance centrale : $\mu_0=0$				
Test	Statistique		p Value	
t de Student	t	6.749558	Pr > t	<.0001
Signe	M	32	Pr >= M	<.0001
Rang signé	S	1040	Pr >= S	<.0001

Tests de normalité				
Test	Statistique		p Value	
Shapiro-Wilk	W	0.709783	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.245451	Pr > D	<0.0100
Cramer-von Mises	W-Sq	1.648859	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	8.996961	Pr > A-Sq	<0.0050

Tableau 3 : estimation des effets fixes avec GLMMIX

Solutions for Fixed Effects								
Effet	bloc	T	H	Valeur estimée	Erreur type	DDL	Valeur du test t	Pr > t
Intercept				0.7873	0.6804	14	1.16	0.2666
bloc	A			-0.3830	0.3201	14	-1.20	0.2514
bloc	B			-0.01571	0.3050	14	-0.05	0.9596
bloc	C			0
T		R1		0.7778	0.8151	14	0.95	0.3562
T		R2		-0.4779	1.0213	14	-0.47	0.6471
T		R3		0.5430	0.8441	14	0.64	0.5305
T		R4		1.0112	0.7948	14	1.27	0.2240
T		R5		0.03345	0.9253	14	0.04	0.9717
T		RLH		0.4436	0.8481	14	0.52	0.6091
T		RS		-0.00584	0.9269	14	-0.01	0.9951
T		RSDH		0
H			0-10 cm	1.8192	0.6462	16	2.81	0.0125
H			10-20 cm	0
T*H		R1	0-10 cm	-1.0519	0.8127	16	-1.29	0.2139
T*H		R1	10-20 cm	0
T*H		R2	0-10 cm	0.05264	1.0196	16	0.05	0.9595
T*H		R2	10-20 cm	0
T*H		R3	0-10 cm	-1.0307	0.8551	16	-1.21	0.2456
T*H		R3	10-20 cm	0
T*H		R4	0-10 cm	-1.7585	0.8242	16	-2.13	0.0487
T*H		R4	10-20 cm	0
T*H		R5	0-10 cm	-0.02740	0.9148	16	-0.03	0.9765
T*H		R5	10-20 cm	0
T*H		RLH	0-10 cm	-0.6878	0.8383	16	-0.82	0.4241
T*H		RLH	10-20 cm	0
T*H		RS	0-10 cm	-0.5664	0.9379	16	-0.60	0.5544
T*H		RS	10-20 cm	0
T*H		RSDH	0-10 cm	0
T*H		RSDH	10-20 cm	0

Tableau 4 : estimation des effets fixes avec *glmer*

Estimation des effets fixes			
Effet	Valeur estimée	Erreur type	valeur du test t
(Intercept)	0,75650	0,98220	0,77020
blocA	-0,40044	0,50180	-0,79801
blocB	-0,00971	0,48126	-0,02017
blocC	0	.	.
TR1	0,78095	1,18217	0,66061
TR2	-0,52190	1,45690	-0,35823
TR3	0,54014	1,22027	0,44264
TR4	1,01122	1,15597	0,87478
TR5	0,04177	1,32522	0,03152
TRLH	0,41992	1,22724	0,34216
TRS	-0,01633	1,33014	-0,01228
TRSDH	0	.	.
H0-10 cm	1,81916	0,88520	2,05509
H10-20 cm	0	.	.
TR1:H0-10 cm	-1,05191	1,11331	-0,94485
TR1:H10-20 cm	0	.	.
TR2:H0-10 cm	0,05264	1,39816	0,03765
TR2:H10-20 cm	0	.	.
TR3:H0-10 cm	-1,03070	1,17221	-0,87928
TR3:H10-20 cm	0	.	.
TR4:H0-10 cm	-1,75854	1,12956	-1,55683
TR4:H10-20 cm	0	.	.
TR5:H0-10 cm	-0,02740	1,25314	-0,02187
TR5:H10-20 cm	0	.	.
TRLH:H0-10 cm	-0,68775	1,14842	-0,59887
TRLH:H10-20 cm	0	.	.
TRS:H0-10 cm	-0,56640	1,28664	-0,44021
TRS:H10-20 cm	0	.	.
TRSDH:H0-10 cm	0	.	.
TRSDH:H10-20 cm	0	.	.

Tableau 5 : analyse des résidus avec GLIMMIX (excepté le test de Levène obtenu avec R)

Moments			
N	48	Somme des poids	48
Moyenne	-0.0226152	Somme des observations	-1.0855275
Ecart-type	0.56359386	Variance	0.31763804
Skewness	0.58135295	Kurtosis	0.95528257
Somme des carrés non corrigée	14.9535373	Somme des carrés corrigée	14.9289879
Coeff Variation	-2492.1068	Std Error Mean	0.08134777

Mesures statistiques de base			
Tendance centrale		Variabilité	
Moyenne	-0.02262	Ecart-type	0.56359
Médiane	-0.00769	Variance	0.31764
Mode	-1.00000	Intervalle	2.63746
		Ecart interquartile	0.71019

Tests de tendance centrale : $\mu_0=0$				
Test	Statistique		p Value	
t de Student	t	-0.27801	Pr > t	0.7822
Signe	M	0	Pr >= M	1.0000
Rang signé	S	-41	Pr >= S	0.6787

Tests de normalité				
Test	Statistique		p Value	
Shapiro-Wilk	W	0.962654	Pr < W	0.1292
Kolmogorov-Smirnov	D	0.086428	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.040864	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.38652	Pr > A-Sq	>0.2500

Test de Levène		
Df	valeur de F	Pr(> F)
15	0.3669	0.9788

Tableau 6 : Table des comparaisons de moyenne deux à deux (obtenue avec GLIMMIX).

Differences of T*date Least Squares Means								
T	date	T	date	Valeur estimée	Erreur type	DDL	Valeur du test t	Pr > t
cajanus	3	cajanus	4	0.000150	0.09654	144	0.00	0.9988
cajanus	3	cajanus	5	0.008250	0.09654	144	0.09	0.9320
cajanus	3	cajanus	6	0.02460	0.09654	144	0.25	0.7992
cajanus	3	cajanus	7	0.09020	0.09654	144	0.93	0.3517
cajanus	3	cajanus	8	0.2016	0.09654	144	2.09	0.0385
cajanus	3	cajanus	9	0.2930	0.09654	144	3.03	0.0029
cajanus	3	desmodiu	1	-0.3129	0.1122	144	-2.79	0.0060
cajanus	3	desmodiu	2	-0.3704	0.1122	144	-3.30	0.0012
cajanus	3	desmodiu	3	-0.3506	0.1122	144	-3.13	0.0021

On constate donc ici que les différences étaient significatives entre les traitements : Cajanus à la date 3, et desmodium aux dates 1,2,3. De plus, pour un même traitement (cajanus), on remarque que les poids des larves étaient significativement différents entre la date 3, et les dates 8 et 9.

Tableau 7 : Résultats de l'ANOVA des indices de diversité de la macrofaune du sol

		effet étudié	traitement	horizon	bloc	traitement:horizon
Riz	mars	richesse	-	significatif	-	-
		shannon	-	significatif	-	-
		Gini-Simpson	-	significatif	-	-
	août	richesse	-	significatif	-	-
		shannon	significatif	significatif	-	-
		Gini-Simpson	-	significatif	-	-
Maïs	mars	richesse	-	significatif	-	-
		shannon	-	significatif	-	-
		Gini-Simpson	-	significatif	-	-
	août	richesse	-	significatif	-	-
		shannon	-	significatif	significatif	-
		Gini-Simpson	-	significatif	significatif	-
plantes seules	mars	richesse	-	significatif	-	-
		shannon	-	significatif	-	-
		Gini-Simpson	-	significatif	significatif	-
	août	richesse	-	significatif	-	-
		shannon	-	significatif	-	-
		Gini-Simpson	-	significatif	-	-

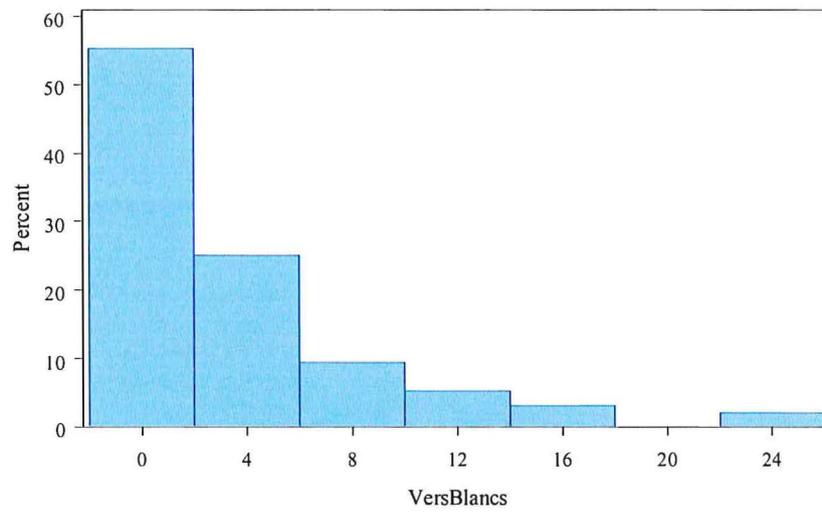
Tableau 8 : Classement des traitements en fonctions de la moyenne de l'estimation du taux de poquets attaqués sur le nombre de poquets restants

taux moyen de poquets attaqués sur les poquets restants par traitement et par date									
Levée		Début tallage		Mi-tallage		Montaison		Epiaison	
RS	2,4%	RSDH	28,5%	RSDH	36,7%	RS	44,5%	R5	33,8%
RSDH	4,1%	R5	35,2%	RS	40,7%	R5	45,6%	RSDH	42,4%
R2	4,5%	RS	38,1%	R5	42,3%	RSDH	48,7%	RS	51,4%
R4	4,9%	R3	39,8%	R3	49,7%	R3	56,1%	R3	52,7%
RLH	5,0%	R4	55,5%	R4	68,6%	R4	65,9%	R1	66,0%
R5	6,9%	R1	57,4%	R1	69,1%	R1	69,8%	R4	67,3%
R3	8,7%	R2	65,8%	R2	84,0%	R2	85,7%	R2	85,6%
R1	9,1%	RLH	76,7%	RLH	85,2%	RLH	92,1%	RLH	93,9%

Graphiques

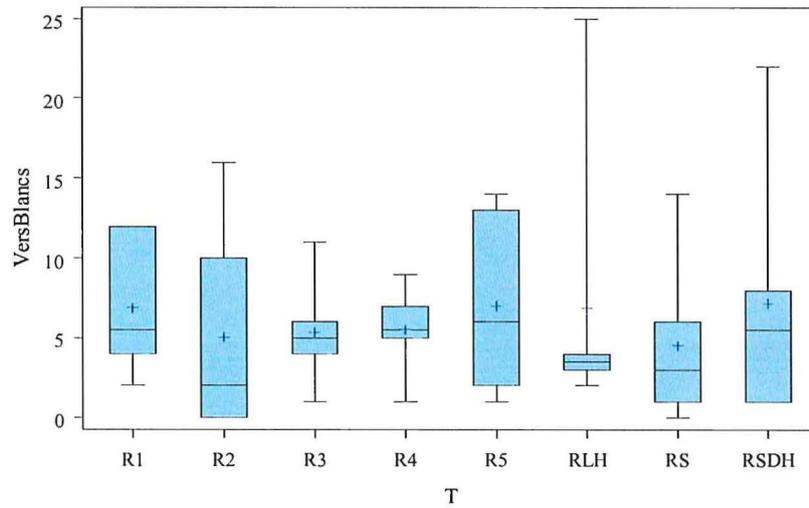
Graphe 1 : histogramme des densités de vers blancs

Densités de vers blancs dans les cultures de riz (mars 2009)

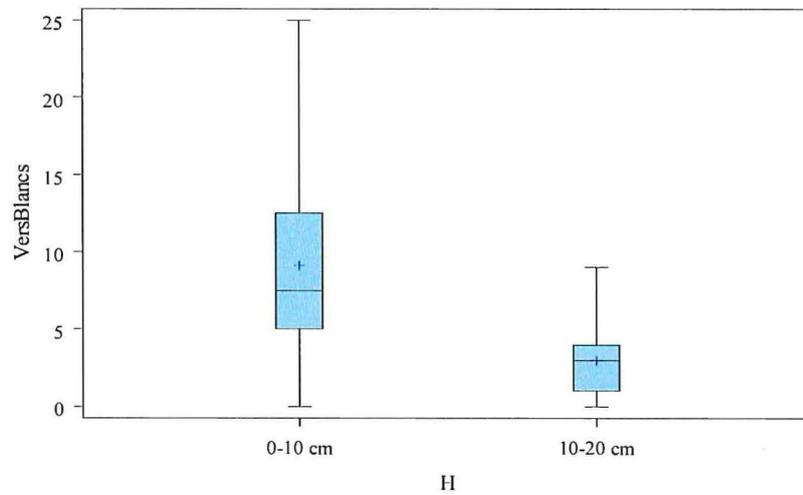


Graphes 2 : boxplot des densités de vers blancs dans les cultures de riz en mars

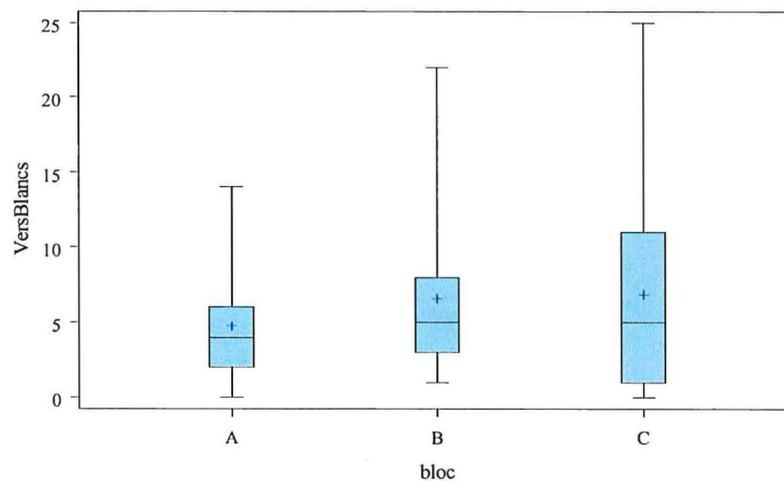
Densités de vers blancs dans les cultures de riz (mars 2009)



Densités de vers blancs dans les cultures de riz (mars 2009)

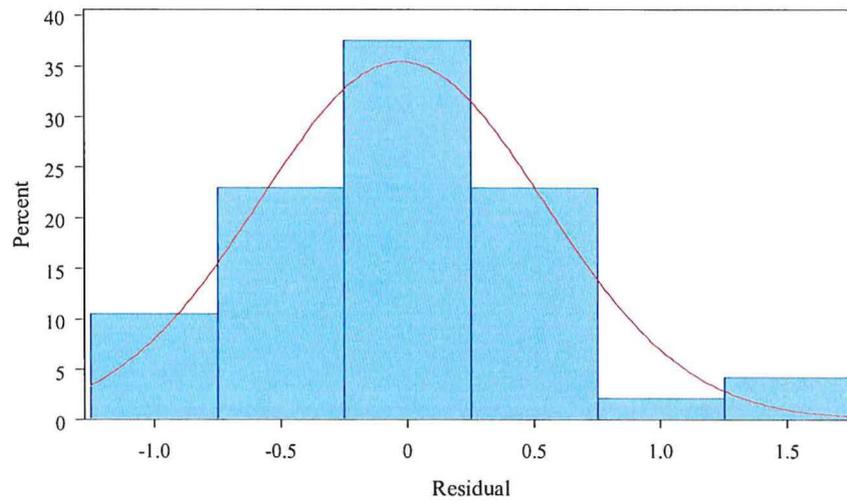


Densités de vers blancs dans les cultures de riz (mars 2009)

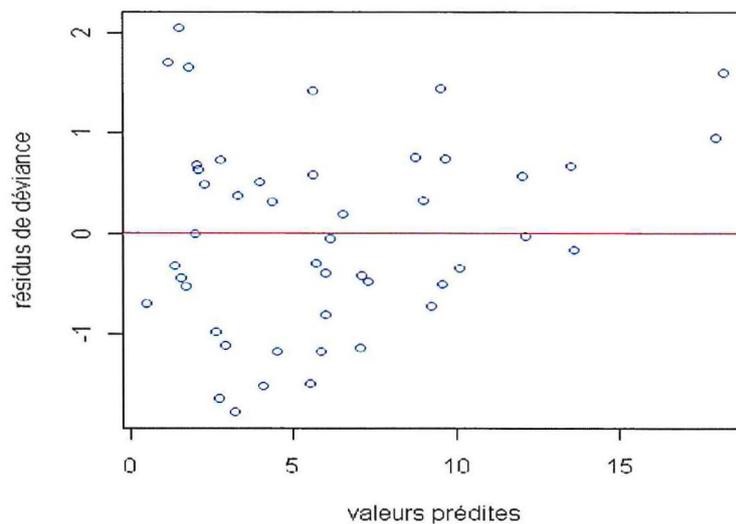
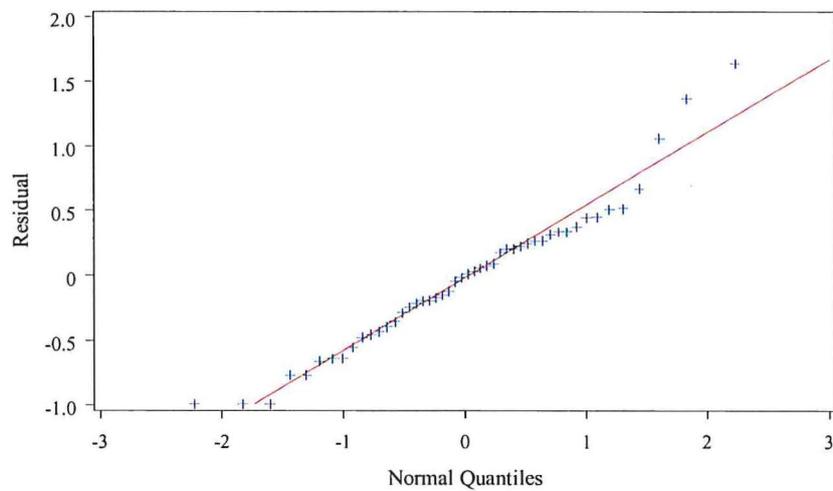


Graphe 3 : graphes des résidus (pour la variable densité de vers blancs dans le dispositif riz (mars))

Densités de vers blancs dans les cultures de riz (mars 2009)



Densités de vers blancs dans les cultures de riz (mars 2009)



Estimations des densités moyennes de vers blancs

- Dispositif : riz pluvial

Mars

Moyenne du nombre de vers blancs par traitement et par horizon			
0-10 cm		10-20cm	
R4	5,6	R2	1,2
RS	6,7	RS	1,9
R3	7,3	RSDH	1,9
R2	7,8	R5	2,0
R1	9,0	RLH	3,0
RLH	9,3	R3	3,3
RSDH	11,9	R1	4,2
R5	11,9	R4	5,3

Août

Moyenne du nombre de vers blancs par traitement et par horizon			
0-10 cm		10-20cm	
R3	0,3	RSDH	0,9
R1	2,1	R1	1,2
RSDH	2,1	R3	1,2
R5	3,0	R2	2,4
R2	3,9	R4	2,4
RS	3,9	RLH	2,4
R4	4,2	R5	2,7
RLH	4,2	RS	2,7

- Dispositif : maïs

Mars

Moyenne du nombre de vers blancs par traitement et par horizon			
0-10 cm		10-20cm	
M5	1,3	M2	0,4
MLH	1,8	MSDH	0,5
M1	2,5	M5	0,5
MSDH	3,1	MLH	0,5
M4	3,5	M1	0,8
MS	3,8	M4	1,1
M2	4,9	MS	1,7
M3	11,6	M3	3,8

Août

Moyenne du nombre de vers blancs par traitement et par horizon			
0-10 cm		10-20cm	
M5	0,4	MLH	0,8
M2	0,5	M4	1,1
MSDH	0,7	MS	1,5
M1	1,0	MSDH	1,6
MS	1,0	M2	2,3
M4	1,1	M5	2,4
M3	1,8	M3	3,4
MLH	2,4	M1	5,4

- Dispositif : plantes seules

Mars

Moyenne du nombre de vers blancs par traitement et par horizon			
0-10 cm		10-20cm	
brachiaria mulato	2,9	cleome viscosa	1,5
radis fourrager	4,6	eleusine+crotalaire	1,8
vesce	5,1	brachiaria mulato	1,9
eleusine+crotalaire	5,1	radis fourrager	2,6
cleome viscosa	6,4	haricot	5,5
haricot	10,4	vesce	6,1

Août

Moyenne du nombre de vers blancs par traitement et par horizon			
0-10 cm		10-20cm	
eleusine+crotalaire	0,7	cleome viscosa	1,6
haricot	0,7	haricot	1,7
cleome viscosa	1,3	eleusine+crotalaire	2,0
brachiaria mulato	2,0	radis fourrager	2,0
vesce	2,0	brachiaria mulato	2,6
radis fourrager	3,0	vesce	3,7

Estimations des indices de diversité moyens de la macrofaune du sol

- Dispositif : riz pluvial

Mars

- Indice de richesse :

Indice de richesse moyen par traitement et par horizon							
Litière		0-10 cm		10-20 cm		20-30 cm	
R2	2,3	R1	3,3	R2	2,0	R1	2,0
RS	2,3	RS	4,7	RS	2,3	R3	2,0
R4	2,7	R2	5,0	R3	2,7	RLH	2,0
R1	3,0	RLH	5,3	R5	2,7	RS	2,0
RSDH	3,0	R4	5,7	RSDH	2,7	R4	2,3
R3	3,3	RSDH	6,0	RLH	3,0	R2	2,7
R5	4,7	R3	6,3	R1	4,0	R5	2,7
RLH	4,7	R5	6,3	R4	4,3	RSDH	3,0

- Indice de Shannon :

Indice de Shannon moyen par traitement et par horizon							
Litière		0-10 cm		10-20 cm		20-30 cm	
RS	0,63	R1	0,82	RS	0,33	R1	0,36
R2	0,66	RLH	0,92	R2	0,50	R4	0,42
R1	0,79	R2	0,97	RSDH	0,64	RS	0,43
R4	0,82	RSDH	1,10	R3	0,74	R3	0,52
RSDH	0,97	R3	1,13	R4	0,85	RLH	0,64
R3	1,09	RS	1,23	R5	0,86	R2	0,86
RLH	1,11	R4	1,32	RLH	0,88	R5	0,90
R5	1,33	R5	1,50	R1	1,07	RSDH	0,95

- Indice de Gini-Simpson :

Indice de Gini-Simpson moyen par traitement et par horizon							
Litière		0-10 cm		10-20 cm		20-30 cm	
RS	0,37	R1	0,44	RS	0,19	R1	0,19
R1	0,44	RLH	0,45	R2	0,31	R4	0,22
R2	0,44	R2	0,48	RSDH	0,38	RS	0,23
R4	0,49	RSDH	0,54	R3	0,42	R3	0,33
RLH	0,56	R3	0,55	R4	0,44	RLH	0,45
RSDH	0,58	R4	0,63	RLH	0,51	RSDH	0,54
R3	0,64	RS	0,64	R5	0,53	R2	0,54
R5	0,68	R5	0,74	R1	0,58	R5	0,57

Août

- Indice de richesse :

Indice de richesse moyen par traitement et par horizon							
Litière		0-10 cm		10-20 cm		20-30 cm	
RLH	2,0	RLH	4,3	R3	2,0	R3	1,0
R1	2,3	R3	5,7	RS	3,3	R2	1,7
R3	2,7	RS	5,7	R4	3,7	R1	2,0
R4	2,7	R1	6,0	RLH	3,7	R4	2,0
R2	3,0	R2	6,0	R1	4,0	R5	3,0
RS	3,0	R4	6,7	R2	4,0	RLH	3,0
RSDH	3,3	R5	6,7	R5	4,3	RS	3,0
R5	4,3	RSDH	7,7	RSDH	5,0	RSDH	4,7

- Indice de Shannon :

Indice de Shannon moyen par traitement et par horizon							
Litière		0-10 cm		10-20 cm		20-30 cm	
RLH	0,56	RLH	1,16	R3	0,59	R3	0,00
R4	0,69	RS	1,33	RLH	0,94	R2	0,35
R3	0,75	R4	1,39	R4	0,94	R4	0,46
R1	0,75	R3	1,43	RS	1,04	R1	0,55
RS	0,85	R2	1,49	R2	1,08	RLH	0,76
R2	0,86	R1	1,53	R1	1,16	RS	0,83
RSDH	1,07	R5	1,69	RSDH	1,21	RSDH	0,91
R5	1,41	RSDH	1,84	R5	1,31	R5	0,92

- Indice de Gini-Simpson :

Indice de Gini-Simpson moyen par traitement et par horizon							
Litière		0-10 cm		10-20 cm		20-30 cm	
RLH	0,36	RLH	0,62	R3	0,40	R3	0,00
R4	0,39	RS	0,65	R4	0,53	R2	0,21
R3	0,43	R4	0,65	RLH	0,55	R4	0,27
RS	0,45	R2	0,71	R2	0,56	R1	0,35
R1	0,49	R3	0,71	RS	0,60	RLH	0,41
R2	0,50	R1	0,73	RSDH	0,62	RSDH	0,43
RSDH	0,62	R5	0,78	R1	0,62	RS	0,44
R5	0,74	RSDH	0,81	R5	0,69	R5	0,56

- Dispositif : mais

Mars

- Indice de richesse :

Indice de richesse moyen par traitement et par horizon							
Litière		0-10 cm		10-20 cm		20-30 cm	
MSDH	1,7	M5	4,0	M2	2,0	M1	1,3
M2	2,0	M4	4,3	M1	3,0	M2	1,3
M4	2,0	MS	4,7	M4	3,7	M3	1,7
M3	2,7	M3	5,0	MLH	3,7	MLH	1,7
M5	2,7	M1	5,3	MSDH	3,7	M5	2,0
MS	2,7	MLH	5,3	M5	4,0	M4	2,3
M1	3,7	M2	5,7	M3	4,7	MS	3,0
MLH	3,7	MSDH	5,7	MS	4,7	MSDH	3,0

- Indice de Shannon :

Indice de Shannon moyen par traitement et par horizon							
Litière		0-10 cm		10-20 cm		20-30 cm	
MSDH	0,42	M5	0,83	M2	0,56	M1	0,21
M4	0,44	MSDH	0,90	M1	0,79	M2	0,21
M2	0,52	MLH	0,95	M5	0,96	M3	0,21
MS	0,72	M4	1,02	MLH	1,00	MLH	0,37
M5	0,73	M3	1,07	M3	1,07	M4	0,41
M3	0,89	MS	1,11	M4	1,13	M5	0,50
M1	1,06	M2	1,33	MSDH	1,14	MS	0,79
MLH	1,17	M1	1,34	MS	1,25	MSDH	1,00

- Indice de Gini-Simpson :

Indice de Gini-Simpson moyen par traitement et par horizon							
Litière		0-10 cm		10-20 cm		20-30 cm	
M4	0,26	MSDH	0,42	M2	0,36	M3	0,11
MSDH	0,30	M5	0,44	M1	0,45	M1	0,15
M2	0,32	MLH	0,46	M5	0,50	M2	0,15
M5	0,42	M3	0,55	MLH	0,56	M4	0,20
MS	0,45	M4	0,55	M3	0,56	MLH	0,22
M3	0,56	MS	0,57	MSDH	0,63	M5	0,31
M1	0,56	M2	0,66	M4	0,63	MS	0,41
MLH	0,65	M1	0,68	MS	0,65	MSDH	0,60

Août

- Indice de richesse :

Indice de richesse moyen par traitement et par horizon							
Litière		0-10 cm		10-20 cm		20-30 cm	
MS	1,7	M4	4,3	MS	1,7	M2	1,3
M3	2,3	MS	4,3	M3	2,0	M5	1,7
MSDH	2,3	M3	4,7	M5	2,0	MSDH	1,7
MLH	2,7	M1	5,6	M1	2,3	M1	2,3
M1	3,0	MLH	5,6	MLH	3,0	MLH	2,7
M4	3,3	MSDH	5,6	MSDH	3,0	M3	3,0
M2	4,3	M2	6,3	M2	4,0	MS	3,3
M5	5,0	M5	6,3	M4	4,0	M4	4,0

- Indice de Shannon :

Indice de Shannon moyen par traitement et par horizon							
Litière		0-10 cm		10-20 cm		20-30 cm	
MS	0,37	MSDH	1,20	M3	0,40	M2	0,23
MSDH	0,52	M4	1,28	M5	0,42	M5	0,35
M3	0,64	M5	1,31	MS	0,42	MSDH	0,37
MLH	0,72	M3	1,32	M1	0,66	M3	0,71
M1	0,79	MS	1,34	MSDH	0,70	M1	0,73
M4	0,90	M1	1,53	MLH	0,87	MS	0,84
M2	1,03	MLH	1,59	M2	1,13	MLH	0,90
M5	1,40	M2	1,77	M4	1,20	M4	1,03

- Indice de Gini-Simpson :

Indice de Gini-Simpson moyen par traitement et par horizon							
Litière		0-10 cm		10-20 cm		20-30 cm	
MS	0,22	MSDH	0,57	M3	0,22	M2	0,17
MSDH	0,28	M5	0,61	M5	0,24	M5	0,21
M3	0,39	M3	0,67	MS	0,29	MSDH	0,25
M1	0,43	M4	0,68	MSDH	0,38	M3	0,36
MLH	0,43	MS	0,71	M1	0,42	MS	0,47
M4	0,51	M1	0,74	MLH	0,52	M1	0,48
M2	0,53	MLH	0,76	M2	0,59	M4	0,56
M5	0,69	M2	0,82	M4	0,66	MLH	0,57

- Dispositif : plantes seules

Mars

- Indice de richesse :

Indice de richesse moyen par traitement et par horizon							
Litière		0-10 cm		10-20 cm		20-30 cm	
radis fourrager	2,7	radis fourrager	5,3	eleusine+crotalaire	2,7	brachiaria mulato	2,3
cleome viscosa	3,0	vesce	6,0	vesce	2,7	vesce	2,3
haricot	3,0	brachiaria mulato	6,7	radis fourrager	3,3	cleome viscosa	2,7
eleusine+crotalaire	3,3	eleusine+crotalaire	7,3	brachiaria mulato	4,0	eleusine+crotalaire	2,7
brachiaria mulato	4,0	cleome viscosa	8,0	cleome viscosa	4,3	haricot	3,3
vesce	4,0	haricot	8,3	haricot	4,7	radis fourrager	3,3

- Indice de Shannon :

Indice de Shannon moyen par traitement et par horizon							
Litière		0-10 cm		10-20 cm		20-30 cm	
cleome viscosa	0,47	radis fourrager	1,20	eleusine+crotalaire	0,63	vesce	0,63
radis fourrager	0,54	vesce	1,23	radis fourrager	0,66	brachiaria mulato	0,81
eleusine+crotalaire	0,75	brachiaria mulato	1,25	vesce	0,73	cleome viscosa	0,81
vesce	0,88	haricot	1,50	brachiaria mulato	1,03	eleusine+crotalaire	0,81
haricot	0,92	eleusine+crotalaire	1,51	haricot	1,04	haricot	0,93
brachiaria mulato	1,23	cleome viscosa	1,67	cleome viscosa	1,24	radis fourrager	0,98

- Indice de Gini-Simpson :

Indice de Gini-Simpson moyen par traitement et par horizon							
Litière		0-10 cm		10-20 cm		20-30 cm	
cleome viscosa	0,27	brachiaria mulato	0,58	eleusine+crotalaire	0,33	vesce	0,37
radis fourrager	0,29	vesce	0,59	radis fourrager	0,39	eleusine+crotalaire	0,46
eleusine+crotalaire	0,41	radis fourrager	0,61	vesce	0,49	cleome viscosa	0,46
vesce	0,44	eleusine+crotalaire	0,70	brachiaria mulato	0,53	radis fourrager	0,51
haricot	0,54	haricot	0,71	haricot	0,54	haricot	0,52
brachiaria mulato	0,68	cleome viscosa	0,77	cleome viscosa	0,65	brachiaria mulato	0,54

Août

- Indice de richesse :

Indice de richesse moyen par traitement et par horizon							
Litière		0-10 cm		10-20 cm		20-30 cm	
haricot	1,66	haricot	4,64	eleusine+crotalaire	2,32	cleome viscosa	1,66
cleome viscosa	2,32	cleome viscosa	5,30	brachiaria mulato	2,65	haricot	1,99
radis fourrager	2,32	brachiaria mulato	5,63	haricot	2,65	brachiaria mulato	2,32
eleusine+crotalaire	2,98	eleusine+crotalaire	5,96	radis fourrager	2,65	eleusine+crotalaire	2,32
vesce	3,31	radis fourrager	6,96	vesce	2,65	radis fourrager	2,98
brachiaria mulato	3,64	vesce	6,96	cleome viscosa	2,98	vesce	2,98

- Indice de Shannon :

Indice de Shannon moyen par traitement et par horizon							
Litière		0-10 cm		10-20 cm		20-30 cm	
haricot	0,34	haricot	1,04	eleusine+crotalaire	0,64	cleome viscosa	0,42
cleome viscosa	0,58	cleome viscosa	1,10	radis fourrager	0,68	haricot	0,46
radis fourrager	0,70	brachiaria mulato	1,48	cleome viscosa	0,76	brachiaria mulato	0,68
eleusine+crotalaire	0,80	eleusine+crotalaire	1,51	brachiaria mulato	0,80	eleusine+crotalaire	0,70
vesce	0,92	vesce	1,70	vesce	0,84	vesce	0,82
brachiaria mulato	1,10	radis fourrager	1,75	haricot	0,86	radis fourrager	1,07

- Indice de Gini-Simpson :

Indice de Gini-Simpson moyen par traitement et par horizon							
Litière		0-10 cm		10-20 cm		20-30 cm	
haricot	0,22	haricot	0,52	eleusine+crotalaire	0,38	haricot	0,25
cleome viscosa	0,32	cleome viscosa	0,54	radis fourrager	0,38	cleome viscosa	0,29
radis fourrager	0,45	eleusine+crotalaire	0,70	cleome viscosa	0,45	brachiaria mulato	0,42
eleusine+crotalaire	0,46	brachiaria mulato	0,71	brachiaria mulato	0,49	eleusine+crotalaire	0,44
vesce	0,55	vesce	0,78	vesce	0,52	vesce	0,45
brachiaria mulato	0,61	radis fourrager	0,79	haricot	0,53	radis fourrager	0,65

Estimations des indices de diversité moyens de la macrofaune des Pitfall trap par traitement

- Dispositif : riz pluvial (août)

Indice de richesse		Indice de Shannon		Indice de Gini-Simpson	
traitement	Moyenne	traitement	Moyenne	traitement	Moyenne
R2	4,3	RLH	0,60	RLH	0,30
RLH	4,3	R3	0,70	R3	0,32
RS	4,6	R4	0,81	RS	0,39
R5	5,0	RS	0,81	R2	0,41
R3	5,3	R5	0,84	R4	0,42
R4	5,3	R2	0,86	R5	0,45
RSDH	6,0	R1	1,04	R1	0,47
R1	6,3	RSDH	1,10	RSDH	0,51

- Dispositif : maïs (août)

Indice de richesse		Indice de Shannon		Indice de Gini-Simpson	
traitement	Moyenne	traitement	Moyenne	traitement	Moyenne
MLH	4,0	M3	0,86	MSDH	0,42
MS	4,0	M2	0,87	M2	0,46
M3	4,3	MSDH	0,89	M3	0,46
M4	4,3	M1	0,98	M1	0,49
M2	4,7	MLH	1,00	M5	0,51
M1	5,0	M4	1,00	MLH	0,53
M5	5,7	M5	1,03	M4	0,54
MSDH	6,0	MS	1,15	MS	0,62

- Dispositif : plantes seules (août)

Indice de richesse		Indice de Shannon		Indice de Gini-Simpson	
traitement	Moyenne	traitement	Moyenne	traitement	Moyenne
haricot	4,7	cleome viscosa	0,35	cleome viscosa	0,13
brachiaria mulato	5,0	vesce	0,46	vesce	0,22
vesce	5,3	haricot	0,62	haricot	0,30
cleome viscosa	6,3	eleusine+crotalaire	0,78	eleusine+crotalaire	0,33
eleusine+crotalaire	6,7	radis fourrager	1,14	radis fourrager	0,50
radis fourrager	7,0	brachiaria mulato	1,14	brachiaria mulato	0,56