



HAL
open science

**Réponses des prairies aux variations du pH du sol :
approche modélisation Rapport de stage de 1 ère année
de master Mathématiques Appliquées, mention
Statistiques**

Sarah Teixeira

► **To cite this version:**

Sarah Teixeira. Réponses des prairies aux variations du pH du sol : approche modélisation Rapport de stage de 1 ère année de master Mathématiques Appliquées, mention Statistiques. Sciences de l'environnement. 2020. hal-04068903

HAL Id: hal-04068903

<https://hal.inrae.fr/hal-04068903>

Submitted on 14 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Clermont Auvergne
UFR de Mathématiques



Sarah Teixeira

Réponses des prairies aux variations du pH du sol : approche
modélisation

Rapport de stage de 1^{ère} année de master Mathématiques Appliquées,
mention Statistiques

Du 4 mai 2020 au 28 août 2020



UREP Unité mixte de Recherche sur l'Ecosystème Prairial

Raphaël Martin - Encadrant principal
Juliette Bloor - Maître de stage
Katja Klumpp - Maître de stage
Norddine Azzaoui - Professeur référent du master

Remerciements

Je tiens à remercier toutes les personnes qui m'ont permis de réaliser avec succès ce stage et qui m'ont aidé lors de la rédaction de ce rapport.

J'adresse en particulier mes remerciements à mes maîtres de stage pour la confiance qu'ils m'ont accordée : M. Raphaël MARTIN, Ingénieur de Recherche - Modélisation, pour le partage de son expertise au quotidien, Mme Juliette BLOOR, Chargée de Recherche, pour son accueil et ses précieux conseils ainsi que Mme Katja KLUMP, Ingénieure de Recherche, pour sa coopération et le temps qu'elle m'a accordé.

Je remercie également toute l'équipe INRAe pour leur amabilité et leur esprit d'équipe.

Enfin je remercie mes professeurs de l'UCA pour leur disponibilité et leur implication dans mon apprentissage tout au long de l'année, en particulier M. Nourddine AZZAOUÏ, professeur référant de ce stage.

Résumé

La montée en puissance de l'agroécologie suscite un regain d'intérêt sur le développement de pratiques agricoles durables. Dans les agroécosystèmes au climat tempéré tels que les parcelles labourées, il est connu que le pH des sols joue un rôle primordial sur leur fertilité et la productivité de leur couvert végétal. Cependant, en raison d'un manque de documentation, aucune conclusion similaire n'a pu être tirée concernant l'effet du pH sur les prairies permanentes. Ainsi les modèles de simulation qui sont fréquemment utilisés pour représenter le fonctionnement des prairies n'intègrent que rarement ces effets. C'est pourquoi, l'objectif de cette étude est d'effectuer une analyse de sensibilité sur un modèle largement utilisé pour simuler des prairies permanentes, PaSim, afin d'étudier l'impact d'une évolution de pH et ainsi voir si ce dernier est en capacité de reproduire les observations en situation. De plus, des méthodes de sensibilité d'analyse seront mises en œuvre afin de mettre en évidence que l'effet du pH sur les émissions de N_2O est bien pris en compte par le modèle PaSim mais ce n'est pas le cas de son impact sur le rendement.

Table des matières

Introduction	6
I. Préambule	7
1) Présentation de la structure d'accueil	7
2) Importance des prairies permanentes	8
3) Impact du pH sur les émissions de gaz à effet de serre et le rendement en prairie	9
4) Intérêt des modèles mathématiques pour étudier des systèmes biologiques complexes	12
5) Description du modèle PaSim	13
II. Tests d'un gradient de pH sur le modèle PaSim	15
1. Lancement des simulations	15
2. Présentation des méthodes utilisées	16
➤ ANOVA	16
➤ Analyse graphique	16
➤ AFC	16
➤ Régression de la biomasse produite (le rendement) en fonction du pH	18
➤ T-test	19
3. Application	19
➤ Significativité des différences des variables de sorties en fonction du pH (ANOVA)	19
➤ Comparaison des résultats avec la littérature scientifique (analyse graphique et AFC)	21
Analyse graphique	21
Variable fnh3_nsoil – Volatilisation de l'ammoniac	21
Variable nnit - Quantité de nitrate présent dans le sol	22
Variable fn2osoilatm – Emission de N ₂ O	23
Variable wnorganic – Quantité de nitrogène organique présent dans le sol	23

AFC	24
➤ Etude sur l'effet du pH sur le rendement (Régression)	25
➤ Comparaison du rendement avec les données observées du terrain (T-test)	26
4. Discussion	27
III. Analyse de sensibilité du modèle PaSim	30
1. Lancement des simulations	30
2. Présentation des méthodes utilisées	32
➤ Méthode de Monte Carlo	33
➤ Représentation graphique	33
➤ Méthode de Saltelli	33
3. Application	34
➤ Méthode de Monte Carlo	34
➤ Méthode de Saltelli	36
4. Discussion	38
Conclusion	39
Bibliographie	40
Webographie	40
Annexes	41

Introduction

A l'heure où le développement durable est de mise, l'utilisation de prairies permanentes permet de concevoir des agrosystèmes permettant de concilier des objectifs économiques et environnementaux. En travaillant sur l'agroécologie de l'écosystème des prairies dans un contexte de changement global, l'Unité mixte de Recherche sur l'Ecosystème Prairial (UREP) d'INRAe, l'Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement, s'inscrit pleinement dans cet objectif. Pour cela, l'unité est impliquée dans de nombreux projets internationaux, européens et nationaux afin d'étudier les services rendus par les prairies selon le mode de gestion appliqué et les conditions pédoclimatiques.

Ce stage s'inscrit dans un projet de recherche européen nommé MAGGE-pH ayant pour objectif d'évaluer les effets du pH du sol sur les émissions de N_2O (un gaz à effet de serre important issu des sols fertilisés). Brièvement, une pratique agricole courante pour les sols acides est un amendement de chaux afin de rehausser le pH et la production végétale. Aujourd'hui, il est connu que le chaulage n'améliore pas seulement la production mais réduit aussi les émissions de N_2O qui sont liées au pH du sol. Les études sur le chaulage se basent généralement sur des expérimentations en laboratoires ou en terres cultivées où l'amendement est appliqué sur un sol nu et est incorporé par le labour. De plus, il y a peu d'études sur les prairies, c'est pourquoi le projet MAGGE-pH cherche à mieux comprendre les mécanismes qui pilotent les impacts du pH sur le sol afin de proposer des stratégies de mitigation pour réduire les émissions N_2O à l'échelle de l'Europe. La modélisation permet d'extrapoler les relations entre les pratiques, le pH et les émissions sur des situations plus larges que l'expérimentation. Les observations sur prairies permanentes réalisées dans le cadre de ce projet vont permettre de voir si un modèle comme PaSim permet de simuler correctement la réponse d'un tel agrosystème à l'augmentation du pH.

Ce rapport présente le contexte du stage, la mise en œuvre de tests avec le modèle PaSim (simulations de deux sites de prairies de moyenne montagne avec un gradient de pH) et l'analyse des sorties du modèle en regard des connaissances issues de la littérature scientifique. Un travail sur les analyses de sensibilité a également été réalisé afin d'appliquer l'une de ces méthodes au modèle PaSim pour connaître sa sensibilité au pH et à d'autres variables d'entrée afin de pouvoir en faire une comparaison.

I. Préambule

1) Présentation de la structure d'accueil

L'Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE) est un organisme de recherche français créé en 2020 suite à la fusion de l'Institut National de la Recherche Agronomique (INRA) et de l'Institut national de Recherche en Sciences et Technologies pour l'Environnement et l'Agriculture (IRSTEA). Cette fusion a été initiée par les ministres de l'Agriculture et de la Recherche en février 2018 afin de devenir un établissement public national à caractère scientifique et technologique placé sous la tutelle conjointe du ministre chargé de la Recherche et du ministre chargé de l'Agriculture.

Avant la fusion, l'INRA était le premier institut de recherche agronomique en Europe et le deuxième dans le monde. Son objectif était de mener des recherches pour une alimentation saine et de qualité, pour une agriculture durable, et pour un environnement préservé et valorisé. Il employait près de 8000 personnes réparties sur 17 centres de recherche en France.

L'INRA a été créé en 1946 suite à la seconde guerre mondiale afin de nourrir la France qui connaissait une pénurie alimentaire sur son territoire. A cette époque, l'agriculture française ne permettait pas de subvenir aux besoins alimentaires du pays. A partir des années 1960 et grâce à l'INRA, la France a pu subvenir aux besoins de sa population. L'INRA s'est alors développé plus localement par la création de pôles régionaux. Dans les années 1970, la France a eu un excédent de ressources alimentaires permettant au pays d'exporter. Ainsi, les missions de l'INRA ont évolué vers l'étude de la qualité alimentaire et de la valeur ajoutée d'un produit. Les recherches s'orientent alors sur la transformation de la matière première agricole en intégrant des recherches en microbiologie pour la qualité des aliments et des sciences de l'ingénierie pour les procédés utilisés. La France devient alors en 1973 l'un des premiers exportateurs alimentaires. Cependant, à cause de la crise énergétique, l'INRA cherche à développer une agriculture autonome et économe pour répondre aux besoins du développement local et de l'environnement.

De son côté l'IRSTEA était, avant la fusion, un établissement public français des recherches en sciences et technologies pour l'environnement et l'agriculture. Ses missions étaient de gérer durablement les eaux et les territoires, de prévenir et d'anticiper les risques naturels et la qualité environnementale. L'institut se composait d'une dizaine de pôles de compétitivité et était engagé dans plus de 300 contrats de recherche avec des industriels et des PME (Petite ou moyenne entreprise). Il employait 1200 personnes et avait pour nom jusqu'en 2012, le centre national du machinisme agricole du génie rural, des eaux et des forêts (Cemagref).

Les centres nationaux d'études techniques et de recherches technologiques pour l'agriculture, les forêts et l'équipement rural (CERAFER) ont été créés en 1971, la dénomination de l'établissement a été modifiée à plusieurs reprises (CTGREF en 1973 puis CEMAGREF en 1982) jusqu'à devenir l'IRSTEA en 2012. Depuis leur création, leur but est de mener des études sur de nombreux domaines tels que l'agriculture de montagne, le suivi des innovations techniques ou les problèmes dus à l'utilisation ou la maîtrise de l'eau. En 1986, il devient un établissement public à caractère scientifique et technologique sous la double tutelle des ministères de l'agriculture et de la recherche. Puis en 2010, lors de la création de l'alliance nationale de recherche pour l'environnement, le centre est devenu le porte-parole de la recherche environnementale. Pendant 2 ans, l'IRSTEA a présidé cette alliance qui avait pour but de donner sa place à la recherche dans le débat public sur les défis environnementaux.

Aujourd'hui, INRAe est composé de 18 centres de recherche dispersés dans toute la France, de 14 départements scientifiques spécialisés dans un domaine précis et de 166 projets de recherche européens. Il a pour objectif de réaliser, d'organiser et de coordonner, à son initiative ou à la demande de l'État, tous travaux de recherche scientifique et technologique dans les domaines de l'agriculture, de l'alimentation, de la forêt, de l'environnement, de l'eau, de la biodiversité, de la bioéconomie, de l'économie circulaire, de la gestion durable des territoires et des risques dans les champs afin d'atteindre les objectifs de développement durable et ainsi devenir un leader mondial de la recherche. L'institut cherche à répondre à de nombreux enjeux sociétaux tels qu'en particulier la sécurité alimentaire et nutritionnelle, la transition des agriculteurs vers des pratiques agricoles plus durables ou encore la gestion des ressources naturelles.

L'Unité mixte de Recherche sur l'Ecosystème Prairial (UREP) est une unité rattachée au département écologie et diversité (ECODIV). Elle se focalise sur l'étude de l'agroécologie de l'écosystème des prairies dans un contexte de changement global (climatiques, pratiques de gestion, ...). L'unité est présente dans des projets internationaux et européens dans de nombreux domaines de recherche tels que le bilan des gaz à effet de serre, la séquestration de carbone, l'assemblage des communautés végétales, les interactions entre les plantes, les sols et les animaux, appelées interactions biotiques et leurs conséquences sur le fonctionnement des prairies.

2) Importance des prairies permanentes

Une prairie est définie par l'UNESCO comme une surface couverte par des plantes herbacées avec moins de 10 % d'arbres ou d'arbustes. Elles représentent plus de 40% de la surface émergée du globe et sont présentes dans la quasi intégralité de l'Europe. Au cours des dernières décennies elles ont été délaissées à cause de l'intensification de l'utilisation des surfaces. Cependant il existe peu d'études sur ces chiffres, les informations sont fragmentaires, rares et imprécises et de nouveaux Etats intègrent l'Union Européenne modifiant ainsi les statistiques étudiées. (J.-L. Peyraud, A. Peeters et A. De Vlieghe, 2012)

Une prairie permanente est définie en Europe depuis 2004 comme une surface utilisée pour la production de plantes herbacées, ressemée naturellement ou cultivée (semée) mais qui n'est pas retournée pendant au moins 5 ans. Ces surfaces sont gérées par des pratiques agricoles telle que la fertilisation, le pâturage et la fauche qui sont indispensables pour sa pérennité. Elles jouent un rôle important pour l'alimentation des (grands) herbivores, la préservation des sols par l'érosion et l'épuration de l'eau, la régularisation du climat par la séquestration du carbone ainsi que l'esthétisme des paysages : ce sont les services écosystémiques. Les agronomes et les écologues utilisent entre autres l'abondance et la diversité des espèces afin de caractériser le potentiel agronomique et environnemental de ces prairies, en plus d'indicateurs plus classiques que sont le rendement, la qualité du fourrage, ... Bien que généralement utilisées pour l'alimentation animale, l'intégration des prairies permanentes dans l'alimentation est une des briques majeures pour la production de produits AOP, la qualité des laits, fromages et viandes étant influencées par les vertus des prairies permanentes. Un tiers de la surface agricole en est ainsi recouverte. De plus, en montagne, en raison des fortes contraintes de pente, de sol ou d'hydromorphie, elles représentent l'essentiel de l'alimentation des herbivores tandis qu'en plaine, leurs présences est moindre puisque les surfaces sont plus aptes à accueillir des cultures plus conventionnelles et souvent plus productives (maïs, blé, ...). (J.-L. Peyraud, A. Peeters et A. De Vlieghe, 2012)

Auparavant, les agroécosystèmes prairiaux n'étaient qu'un support de productivité. L'objectif était essentiellement de produire du fourrage pour nourrir le bétail. Mais il y a plusieurs années, une prise de conscience (inter)nationale sur la multifonctionnalité et les services rendus par les prairies a vu le jour et a permis de mettre en évidence l'intérêt de ces écosystèmes pour la planète. La notion de service écosystémique émerge dans le cadre du Millenium Ecosystem Assessment, publié en 2005 puis a été réutilisée et définie par Fisher et Turner en 2008 comme étant l'ensemble des propriétés des écosystèmes permettant de produire le bien-être de l'homme ou de la nature et se divise en trois grandes catégories :

- Les services produits contribuent directement à la production agricole ou permettent de fournir de la nourriture aux animaux.
- Les services environnementaux contribuent quant à eux indirectement à la production agricole, ils permettent de maintenir la qualité de l'environnement par, entre autres, la séquestration du carbone ou bien le maintien de la qualité des eaux (lixiviation de nitrate) et de l'air (émissions de CO₂ et de protoxyde d'azote, N₂O). Cela est permis par l'utilisation d'une couverture végétale composée d'un ensemble d'espèces ayant des effets complémentaires entre elles.
- Les services intrants fournissent des ressources, maintiennent les supports de production agricoles et assurent la régulation biologique (c'est-à-dire entre les espèces) et celle des ressources contribuant au fonctionnement des cycles. Ils sont subdivisés en de nombreux autres : le service de stabilité structurale et de fertilité des sols, le service de disponibilité et de qualité en eau et en air, le service de contrôle des bio-agresseurs et des invasions biologiques et le service des conservations en espèces. (B. Amiaud et P. Carrère, 2012)

Le défi qui se présente à l'agriculture est de développer des systèmes multifonctionnels et durables à la fois productifs et respectueux de l'environnement en prenant en compte les services écologiques assurés par ces agroécosystèmes (Carrère et al., 2012). Les prairies permanentes ont un bon potentiel de séquestration de carbone qui peut permettre de compenser une partie des émissions de gaz à effet de serre, elles protègent le sol contre l'érosion, améliorent sa fertilité, participent à la régulation des flux de nutriments et à la recharge en eau des nappes phréatiques. Cependant, les pratiques de gestion des prairies peuvent modifier le niveau de services qui en découlent, par exemple, un apport d'engrais pour augmenter la production peut augmenter les émissions de gaz à effet de serre (CO₂, N₂O). C'est pourquoi il est intéressant de développer et de mettre en place des pratiques de gestion qui favorisent la fourniture de services écosystémiques et qui minimisent les effets négatifs sur l'environnement.

3) Impact du pH sur les émissions de gaz à effet de serre et le rendement en prairie

Le pH est au centre du sujet de ce stage, il est donc primordial d'en donner une définition succincte. Le pH est l'abréviation du potentiel hydrogène soit un indicateur d'acidité lorsque sa valeur est inférieure à 7 ou d'alcalinité lorsqu'elle est supérieure à 7 (Soren Sorensen, 1893). Cette valeur n'a pas d'unité et est définie par la quantité d'ions hydronium H₃O⁺ présents dans un milieu.

Il est intéressant d'étudier, en milieu prairial, les conséquences d'une modification du pH du sol sur les cycles du carbone (C) et de l'azote (N) et donc sur les émissions de gaz à effet de serre. Ce

sujet est d'actualité puisque ces émissions sont à l'origine, au moins en partie, des changements climatiques récents. Si dans le cadre de l'agriculture, cela s'est dans un premier temps concentré sur le dioxyde de carbone et le méthane, aujourd'hui, réduire les émissions de N₂O est également un enjeu crucial.

Le cycle d'azote est étroitement lié au cycle de carbone. En ce qui concerne le cycle de carbone, les prairies permanentes permettent, généralement, de stocker du carbone dans le sol, jouant ainsi le rôle de puit de carbone même si bien sûr, certaines prairies avec des conditions pédoclimatiques peu favorables combinées à des modes de gestion intenses restent émettrices de CO₂. (D. Vandaele, A. Lebreton et B. Faraco, 2010). Pour limiter les émissions de CO₂, il est essentiel de stocker un maximum de carbone c'est-à-dire d'en conserver plus que ce qui est restitué. Grâce au processus naturel de la photosynthèse, l'herbe des prairies utilise le dioxyde de carbone de l'air (CO₂), l'énergie solaire et l'eau pour pousser. Le carbone s'accumule ainsi dans les tissus végétaux, puis dans le sol sous forme de matière organique quand une partie de la biomasse de la plante est sénescente. De plus, au sein d'une prairie, la respiration des bactéries, des animaux et des végétaux est également à prendre en compte dans l'étude de l'émission de CO₂. C'est pourquoi, les sols des prairies ont un rôle très important dans le stockage de carbone puisque c'est le seul compartiment en capacité de le stocker de manière pérenne. Il est ainsi très important d'éviter de retourner les sols de ces prairies afin d'éviter la libération d'une partie du carbone stocké. (A. Gac, J.-B. Dollé et al., 2010)

La figure 1 décrit les flux de carbone d'une prairie pâturée. L'animal se nourrissant de plantes, le carbone fixé par celles-ci se divise en 3 parties, la première est restituée à l'atmosphère par la respiration des animaux (rejet de CO₂), la seconde l'est par leurs excréments (rejet de méthane) et la dernière retourne au sol par les déjections animales (apport de carbone au sol).

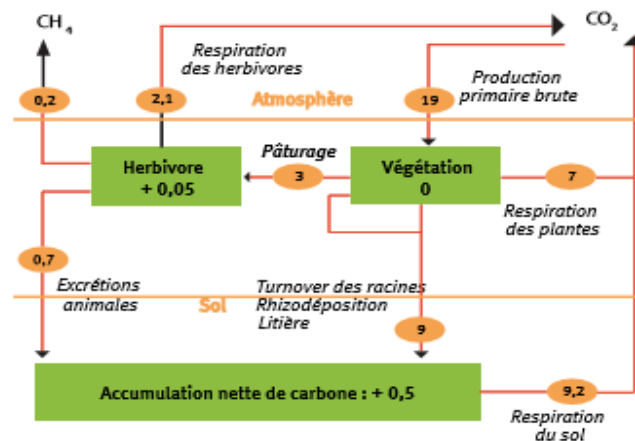


Figure 1 : Flux de carbone d'une prairie pâturée, en tC/ha/an (Soussana et al., 2004)

Du côté du cycle de l'azote, le protoxyde d'azote issu de la fertilisation azotée (minérale et organique) est le troisième GES anthropique en termes de potentiel de réchauffement climatique avec un effet 298 fois plus fort que celui du CO₂. C'est le plus grand contributeur, issu de l'homme, destructeur de la couche d'ozone stratosphérique (A. Viard, C. Hénault et al., 2013) ce qui explique l'intérêt de ce gaz au sein de l'étude de ce stage. La production (et la consommation) de N₂O dans les sols provient des deux processus microbiens principaux présentés par la figure 2 :

- La nitrification qui est l'oxydation biologique de l'engrais comme l'ammonium (NH₄⁺) en nitrite (NO₂⁻) puis en nitrate (NO₃⁻) ;

- La dénitrification qui est un processus respiratoire au cours duquel les formes solubles de l'azote comme le nitrate (NO_3^-) et le nitrite (NO_2^-) sont réduites en composés gazeux (oxyde nitrique NO , protoxyde d'azote N_2O et azote gazeux N_2).

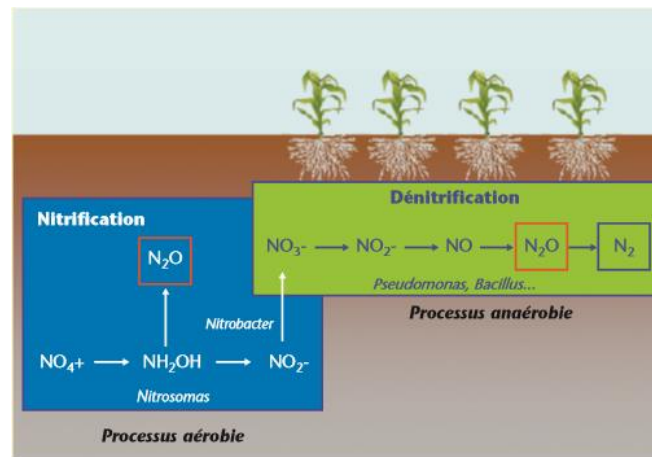


Figure 2 : Processus microbiens à l'origine de la production de N_2O dans les sols (P. Cellier, INRA UMR EGC, 2013)

Pour réduire ces émissions sans pour autant réduire la fertilisation ainsi que la production agricole, certaines études se sont portées sur la dernière étape de la voie de dénitrification dans les sols. Il s'avère en effet que le pH est un levier dans la diffusion de ces gaz. Ainsi, lorsqu'il est inférieur à 6,8, la quantité de N_2O semble être progressivement inhibée. C'est ce qui explique la pratique du chaulage des sols acides jusqu'à la neutralité : cela rend la réduction du N_2O plus efficace et permet de diminuer les émissions de N_2O du sol. (Hénault et al., 2019).

Plus de la moitié des émissions de N_2O est dû à l'implantation d'azote réactif dans le sol par l'utilisation d'engrais synthétiques et la présence de fumier animal. Mais, bien que l'apport d'engrais azotés soit le facteur le plus important expliquant les émissions de N_2O , elles sont également déterminées par le pH du sol qui joue un rôle essentiel. Avec le même apport d'engrais azoté et dans les mêmes conditions, un sol acide émet bien plus de N_2O qu'un sol alcalin.

Les sols fertilisés acides représentent environ 37% des sols français. La technique du chaulage est alors fréquemment utilisée pour remédier à l'acidité d'un sol et ainsi absorber plus d'azote, pour avoir une meilleure biodisponibilité, une meilleure absorption du phosphore et pour améliorer la composition botanique afin de produire plus et mieux. Cette pratique permet de réduire les émissions de N_2O de 9 à 62% d'après une étude sur des sols forestiers allemands menée par Borken et Brumme en 1997. Elle pourrait donc potentiellement réduire les émissions totales de N_2O françaises de près de 15,7% tout en gardant une bonne rentabilité du sol. Elle consiste à apporter un amendement basique calcique dans un sol afin que celui-ci soit moins acide. Mais cette action a directement des conséquences sur la fertilité du milieu. C'est pourquoi, en raison des différents autres impacts possibles du chaulage, il est essentiel de s'intéresser à des études locales qui prennent en compte les aspects agronomiques, environnementaux et économiques car il dépend de l'état initial du sol, de son histoire acido-basique et des pratiques de fertilisation. (B. Fabre et F. Kockmann, 2006)

Il existe différentes méthodes pour estimer des émissions de N_2O : la plus répandue est la méthode de niveau 1 (Tier1) du GIEC (Groupe d'experts intergouvernemental sur l'évolution du climat) qui prend en considération uniquement la quantité d'azote apportée (engrais minéral et fertilisation

organique) à la parcelle comme facteur régissant les émissions de N_2O . Les autres méthodes prennent en compte des processus biochimiques du sol et des plantes des plus simples comme Tier2 du Giec (exemple Henault et al 2005 : « Predicting *in situ* soil N_2O emission using NOE algorithm and soil database ») au plus complexe comme Tier3, qui utilise la modélisation mécaniste (exemple PaSim).

4) Intérêt des modèles mathématiques pour étudier des systèmes biologiques complexes

Mieux comprendre l'impact du pH sur le fonctionnement de l'écosystème prairial est un préalable nécessaire à la mise en place de pratiques agricoles pour une gestion durable des prairies sur sol acide. Cependant les chercheurs sont confrontés à des difficultés pour étudier ce phénomène en situations réelles et ce sur une large gamme de pH du fait de problématiques logistiques, financières, ... C'est pourquoi il est utile de se tourner vers la modélisation qui offre aux biologistes et aux écologistes un moyen plus aisé pour étudier des systèmes et tester des hypothèses en s'affranchissant des difficultés logistiques. Les modèles permettent de tester des théories générales sur la façon dont ces systèmes fonctionnent et la simulation de systèmes avec des modèles aide à identifier les besoins en données et les lacunes dans les connaissances. (Jackson et al, 2000)

Un système peut être considéré comme un ensemble d'entités interagissant entre elles, avec leur environnement et avec les phénomènes créés par ces interactions. Ces entités peuvent être par exemple, dans le cas d'écosystème, un animal, un être humain, un organe ou encore une cellule.

Modéliser un système consiste à utiliser un formalisme (mathématique, graphique, ...) pour expliciter les entités et leurs interactions. Dans le cas d'un système dynamique, cela va inclure un certain nombre de règles d'évolution. Il est alors possible de qualifier de complexe un système lorsque son résultat ne peut pas se déduire uniquement par la connaissance des dites règles (propriétés émergentes).

Bien qu'il n'existe pas de définition universelle, un système complexe présente un certain nombre de caractéristiques :

- Le graphe d'interaction n'est pas trivial (des liens sont privilégiés entre certaines entités) ;
- Les interactions sont locales (pas ou peu d'organisation centrale) ;
- Des boucles de rétroaction sont présentes (l'état d'une entité a une influence sur le futur de cette entité) ;
- La causalité est circulaire (un changement d'état d'une entité va entraîner le changement d'état d'autres entités qui vont, in fine, entraîner un changement d'état chez l'entité initiale. Il est possible par exemple de penser à la notion de cercle vicieux/vertueux.).

Il est possible de constater que généralement les systèmes complexes peuvent être décomposés en groupes d'entités fortement liés. Ainsi, chaque entité interagit localement avec un nombre limité d'autres entités. En extrapolant cela, il est alors possible de décomposer chaque système complexe comme une collection de sous-systèmes complexes en interaction. Une société peut ainsi être vue comme un système composé d'individus en interaction, chaque individu peut être vu comme un système composé d'organes en interaction, chaque organe peut également être vu comme un système composé de cellules en interaction, ...

Les boucles de rétroaction (interaction réflexive soit directe, soit via une chaîne d'interactions) sont une des raisons de la non-linéarité du comportement des systèmes complexes. Ainsi il est possible d'observer dans les cas les plus simples des effets d'emballement, de relaxation ou encore d'oscillation

autour d'un point fixe. Cela implique des comportements difficilement prédictibles une fois que le système intègre de nombreux individus, il est alors compliqué de les étudier par une approche mathématique, c'est pourquoi le passage à l'outil informatique est essentiel.

Dans le cadre du sujet de ce stage au sein de l'UREP, étant donné que les prairies permanentes soient des milieux plurispécifiques avec des interactions entre l'atmosphère, les plantes, le sol et les animaux ainsi que des flux de matières entre ces différentes composantes, un recours à l'outil informatique a été nécessaire. Cela permet d'inclure des contraintes, des variables, des règles ou encore des usages afin que chaque fonction présente dans le modèle puisse être recensée, caractérisée, ordonnée, hiérarchisée et valorisée.

5) Description du modèle PaSim

PaSim (Pasture Simulation Model) est un modèle mécaniste biogéochimique d'écosystème prairial permettant de simuler et de prédire le bilan des flux de carbone, d'azote, d'eau et d'énergie à l'interface entre le sol, la végétation, les animaux et l'atmosphère (Riedo et al. 1998). Le modèle est basé sur le « Hurley pasture model » créé par Thornley en 1988 puis il a été modifié par Riedo et al (Agroscope, suisse) entre 1997 et 2002. Depuis 2002, il est pris en charge par l'UREP de Clermont-Ferrand.

Aujourd'hui, le modèle simule le fonctionnement d'une parcelle de prairie permanente en considérant les interactions entre animaux, plante, sol et atmosphère, dont les flux des principaux GES (méthane, dioxyde de carbone, protoxyde d'azote). PaSim est composé de nombreux (sous-)modules représentant les principaux processus comme la végétation, le microclimat, la biologie et la physique du sol, la gestion agricole et l'animal (figure 3). Le modèle tourne à pas de temps horaire ou journalier. Compte tenu des nombreuses fonctionnalités le nombre minimum de données à entrer est 40 et le nombre maximum est 250. Les 40 paramètres principaux renseignés en entrée du modèle sont :

- Le climat : les facteurs physico-chimiques de l'environnement (température, pluies, humidité, rayonnement, vitesse du vent) et la composition en gaz de l'atmosphère (CO₂).
- Le sol : sa profondeur, sa texture, sa densité apparente et son profil hydrique.
- La végétation : le type de prairie (permanente ou temporaire) ainsi que les paramètres principaux permettant de décrire les caractéristiques du couvert moyen (proportion de légumineuses, hauteur du couvert, etc.)
- Les herbivores (vaches allaitantes et leurs veaux, vaches laitières, génisses ou moutons) : leur état, leur poids et leur âge à la mise à l'herbe ainsi que leur potentiel de lactation dans le cas des vaches laitières.
- La gestion de la parcelle : périodes de pâturage (date, durée et chargement), fauches (date) et fertilisation minérale et/ou organique (dates, nature et quantité).

Le modèle a jusqu'à 480 sorties, les principales sorties (figure 3) comprennent les différents flux de GES, de carbone, d'azote ou encore d'eau mais aussi des variables d'états comme le rendement suite aux fauches, l'ingéré suite au pâturage, le stockage de carbone sous forme de carbone organique, ... Enfin un module d'optimisation permet de calculer les dates optimales de fauches, de pâturages, de fertilisations ainsi que les chargements et quantités associés.

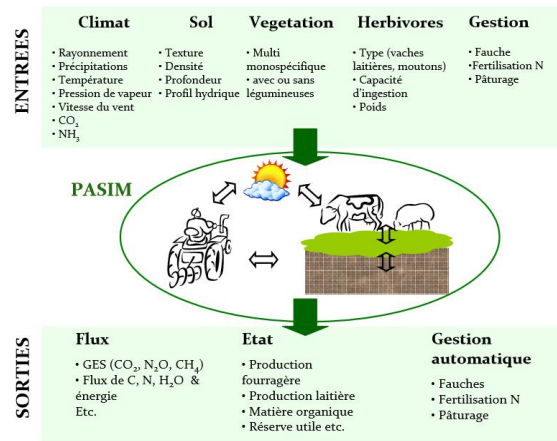


Figure 3 : Présentation de la structure du modèle PaSim

L'ensemble des sorties sont exprimés en m². La durée des simulations se déroulent sur plusieurs années, permettant ainsi d'intégrer l'effet de la variabilité climatique aux prédictions du modèle. Il permet ainsi de simuler des situations sous climat futur, fournissant des pistes pour s'adapter au changement climatique et atténuer leurs effets.

Ce modèle a toutefois de nombreuses limites. Par exemple, les émissions de CH₄ issues du sol ne sont pas modélisées, de même que la végétation de type tropicale ou encore les effets des éventuels maladies, parasites ou ravageurs.

Ses points forts reposent sur la simulation du bilan énergétique, l'approvisionnement en fourrage, la performance des ruminants au pâturage, le potentiel de séquestration du carbone dans les sols de prairie, les émissions de GES au pâturage et les effets climatiques sur le fonctionnement des écosystèmes prairiaux.

Bien que le pH intervienne à plusieurs endroits dans le modèle PaSim, il n'y a jamais eu aucun test explicite pour explorer les réponses et la sensibilité de PaSim à ce paramètre. C'est pourquoi, dans le cadre de ce stage, l'objectif est de quantifier la capacité du modèle PaSim à reproduire les observations du terrain et évaluer sa sensibilité aux variations de pH.

II. Tests d'un gradient de pH sur le modèle PaSim

1. Lancement des simulations

Afin de tester la capacité de PaSim à reproduire des réponses à la variabilité du pH, le modèle a été paramétré sur deux sites de prairies permanentes de moyennes montagnes avec différents modes de fertilisation et en paramétrant un gradient de pH.

Plus en détail, dans le cadre de ce stage, des parcelles de prairies permanentes de moyennes montagnes du Massif Central (Auvergne, près de Rochefort Montagne) ont été simulées ; une située sur le site de Laqueuille avec une simulation sur une durée de 18 ans (de 2002 à 2019). Cette parcelle est une prairie permanente pâturée avec une exploitation extensive (chargement animal faible, 0,6 UGB/ha/an et pas de fertilisation), et deux parcelles du site de Theix (bloc 1) avec une simulation sur 12 ans (de 2005 à 2017). Ces 2 parcelles sont sous fauche (3 par an) sans fertilisation (fauche-zéro) et avec fertilisation NPK (fauche-NPK, 200 kg N/ha.an).

Le modèle a été paramétré avec des données initiales pour chaque parcelle (annexes 1 à 3). Chaque parcelle a ensuite été simulée en modifiant le pH et le faisant évoluer entre 4 et 8 avec un pas de 0.2, de fait 21 simulations ont été réalisées à pas journalier pour 18 ans sur le site de Laqueuille et 12 ans sur les sites de Theix. Parmi les nombreuses sorties proposées par PaSim un panel de variables supposées pouvoir être impactées par le pH a été sélectionné, les variables de sortie du modèle sont les suivantes :

Production végétale

- wshtotsum : le rendement soit le total de matière sèche de la pousse, kg DM m-2
- intake : la dose journalière des animaux par m², kg DM m-2 d-1
- gpp : la productivité primaire brute (PBP), kg C m-2
- npp : la productivité primaire nette (PPN), kg C m-2
- gsh : le taux de croissance des pousses, kg m-2 j-1
- gr : le taux de croissance des racines, kg m-2 j-1
- lai : l'indice de surface foliaire, m² feuille m-2 sol,

Echange de carbone entre la plante, le sol et l'atmosphère

- RECO : la respiration de l'écosystème, kg C m-2
- NEE : les échanges nets de l'écosystème, kg C/m²
- wcorganic : la quantité de carbone organique total dans le sol, kg C m-2

Echange d'azote entre la plante, le sol et l'atmosphère

- fn2o soilatm : la quantité de N₂O échangée entre le sol et l'atmosphère, kg N m-2 d-1
- leachsum : la lixiviation des nitrates, kg N m-2
- fnh3_nsoil : la volatilisation de l'ammoniac à partir du sol, kg N m-2 d-1
- nammtot : la quantité d'ammonium d'azote total dans le sol, kg N m-2
- nnittot : la quantité de nitrate d'azote total dans le sol, kg N m-2
- namm : la quantité d'ammonium d'azote dans la couche du sol h, kg N m-2
- nnit : la quantité de nitrate d'azote dans la couche du sol h, kg N m-2
- wnorganic : la quantité de nitrogène organique total dans le sol, kg N m-2,

Climat du sol

- drainage : le drainage quotidien, mm d-1
- evapotrans : l'évapotranspiration, mm
- thetas : la teneur volumétrique en eau de la couche de sol h, m³ m⁻³
- tsoil : la température de la couche de sol, k

Pratique agricole

- nanimaltot : le taux de chargement, animal m-2
- methane : les émissions quotidiennes de méthane par herbivore, kg C m-2 j-1

2. Présentation des méthodes utilisées

➤ ANOVA

Le choix des variables à analyser s'est fait grâce à une ANOVA à un facteur qu'est le pH. Au sein de ce stage, les moyennes étudiées sont celles de chaque variable en fonction du niveau de pH sur toutes les années confondues. Ainsi, une p-value significative impliquera que les moyennes entre les différents niveaux de pH sont différentes sur l'ensemble de la simulation étudiée. Ces conclusions sont tirées sous R, grâce aux fonctions `aov()` et `summary()` permettant respectivement de calculer une ANOVA et d'en afficher les principales statistiques. Les moyennes seront ensuite comparées 2 à 2 grâce aux fonctions `TukeyHSD()` (qui crée l'ensemble des intervalles de confiance basé sur la méthode de Tukey "Honest Significant Difference") et `pairwise.t.test()` (qui effectue un t-test pour chaque comparaison de paires de variables).

De par l'utilisation d'un gradient de pH lors de la simulation des données, une ANOVA pourrait ne pas être un choix judicieux et la régression pourrait être à privilégier. Cependant, le résultat attendu est une relation non linéaire entre les variables avec effet de seuil ce qui permet de justifier ce choix. Ainsi, l'ANOVA pourra permettre d'étudier les données simulées avec un gradient de pH sur le site de Laqueuille et ceux de Theix.

➤ Analyse graphique

Visualiser les données peut permettre d'en tirer des tendances. Pour chaque simulation, chaque variable prend 4380 valeurs pour les sites de Theix et 6570 pour celui de Laqueuille, il n'est donc pas judicieux de simplement les représenter au cours du temps. C'est pourquoi, pour visualiser au mieux l'évolution des variables au cours du temps et en fonction de chaque pH, la création de moyenne par année et par pH est nécessaire. Ainsi elles peuvent être représentées sous forme de boîtes à moustache, chaque boîte étant les valeurs d'un certain niveau de pH au cours des 12 ou 18 ans mais ces données seront également représentées par des courbes de différentes couleurs, chaque couleur correspondant aux valeurs prises par une année spécifique.

Les variables à représenter ont été choisies au regard des résultats de l'ANOVA. Afin de pouvoir comparer l'évolution des valeurs en fonction des différents modes de fertilisation, 4 d'entre elles ont été sélectionnées : `nmit`, `fn2osoilatm`, `fnh3_nsoil` et `wnorganic`.

➤ AFC

L'AFC (Analyse Factorielle des Composantes) est une extension de l'Analyse en Composantes Principales (ACP), elle permet d'analyser l'association entre deux variables qualitatives (ou

catégorielles). Dans le cadre de ce stage, les AFC ont pour but d'étudier le fonctionnement global du système et d'en déterminer la part d'inertie expliquée par les niveaux de pH. Cette étude est réalisée sur les trois parcelles afin d'avoir des résultats pour un système d'agriculture extensive, un système sans fertilisation et un avec fertilisation NPK.

Les valeurs des variables étudiées sont triées en 10 classes correspondant aux déciles. Ainsi, les 10% des plus petites valeurs seront égales au 1^e décile, les 10% suivant au deuxième décile, ... et les 10 derniers % à la valeur maximale prise par la colonne. Il est alors possible de créer un tableau de contingence de ces données. Les variables à étudier seront sélectionnées grâce à un test de Khi-Deux pour s'assurer de la dépendance des valeurs prises par les variables étudiées avec le niveau de pH.

Puis lors de l'exécution sous R de l'AFC, une fonction nommée returnAFC (voir figure 4) a été créée permettant de renvoyer l'ensemble des graphiques nécessaires à la bonne réalisation de l'interprétation des résultats. Les paramètres d'entrée de cette fonction sont : la table de contingence de la variable étudiée, la table de données initiales et le numéro de la colonne de la variable étudiée au sein de la table de données. Elle permet ainsi :

- De connaître les différentes valeurs des déciles notées D1, ..., D10 pour plus de lisibilité sur les graphiques ;
- De présenter le pourcentage d'explication contenu dans chaque axe de l'AFC ainsi que la contribution moyenne d'un axe qui peut alors servir de critère de décision sur le nombre d'axe à étudier ;
- De représenter la qualité d'interprétation du modèle à l'aide des \cos^2
- D'afficher la contribution des lignes et des colonnes sur les axes 1 et 2 (en prenant en compte le pourcentage d'information contenu dans chaque axe)
- Et enfin de renvoyer la représentation dite biplot des déciles de la variable

```

returnAFC <- function(contingence, data, var){
  # Initialisation

  AFC <- as.data.frame.matrix(contingence)

  nom <- c()
  for (i in 1:nrow(contingence)){
    nom <- c(nom, paste("D",i))
  }

  row.names(AFC) <- nom
  S <- CA(AFC, graph = F)
  # S=CA(AFC, ncp = 5, graph = F)
  # ncp fait 1 AFC et stocke les coordonees des 5 premieres variables

  n <- sum(contingence)
  nc <- ncol(contingence)
  nl <- nrow(contingence)

  quantile(data[,var], probs = seq(0, 1, 0.1))

  # Pourcentage d explication contenu dans chaque axe
  # ligne rouge = critere de decision

  print(fviz_screplot(S, addlabels = TRUE, ylim = c(0, 100), title = "Pourcentage d explication contenu dans
chaque axe") +
  geom_hline (yintercept = 100*1/min(nl-1,nc-1), linetype = 2, color = "red"))

  print(fviz_ca_row(S, col.row = "cos2", gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"), title = "Represen
tation de la qualite d interpretation du modele"))

  print(fviz_contrib(S, choice = "row", axes = 1:2, top = 10, title = "Contribution des lignes sur 1 axe 1 et 2
(en prenant en compte la contribution des axes)")

  print(fviz_contrib(S, choice = "col", axes = 1:2, top = 10, title = "Contribution des colonnes sur 1 axe 1 et
2 (en prenant en compte la contribution des axes)")

  fviz_ca_biplot(S, repel = TRUE, title = paste0('Représentation biplot ', colnames(data)[var]))
}

```

Figure 4 : Script R fonction AFC

➤ Régression de la biomasse produite (le rendement) en fonction du pH

Une analyse graphique préliminaire a été réalisée sur le rendement du site de Theix avec fertilisation NPK et en l'absence de fertilisation. Les tendances observées étaient contraires à ce qui pouvait être attendu, cependant, une régression du pH sur le rendement par année est tout de même réalisée afin de mieux comprendre ces résultats. Pour cela, la fonction `lm()` est utilisée. La qualité d'ajustement et de prévision du modèle est analysée grâce à la fonction `summary()` et la significativité des interactions entre le pH et le rendement est calculée grâce à la fonction `anova()`. Enfin, cette régression est représentée grâce au script présenté dans la figure 5 (et au package `ggpubr`).

```

reg_npk = lm(wshtotsum[year == 1]~pH[year == 1], data = Theix_npk)
reg_npk

```

```

nbYear = 12
for (i in 1:nbYear){
  titre <- paste0("Rendement en fonction du pH - Année = ",i)
  Rendement = Theix_npk$wshtotsum[Theix_npk$year == i]
  pH = Theix_npk$pH[Theix_npk$year == i]
  reg_npk = lm(wshtotsum[year == i]~pH[year == i], data = Theix_npk)
  print( ggplot(reg_npk, aes_string(x=pH, y = Rendement)) +
  labs(titre = "Régression linéaire", subtitle = titre) +
  geom_point() +
  geom_smooth(method=lm, linetype="dashed", color="red"))
}

```

Figure 5 : Script R Représentation de la régression du rendement (Theix fertilisation NPK)

➤ T-test

Les données réelles des sites de Theix sont connues entre les années 2006 et 2018. Comme les simulations ont été lancées sur les années 2005 à 2016, la comparaison des rendements simulés aux rendements réels pourra se faire sur les années 2006 à 2016. Ces données ont dans un premier temps été converties dans la même unité puis représentées graphiquement avec un gradient de couleurs pour les différentes valeurs prises par le pH grâce à la fonction plot(). (annexe 4)

Puis un t-test a été mis en place sous R avec la fonction t.test(), le but étant de comparer les moyennes des différents échantillons de données. Des tests bilatéraux (alternative= "two.sided") et unilatéral inférieur (alternative= "less") ont été effectués sur les deux parcelles de Theix entre les données réelles et simulées.

3. Application

➤ Significativité des différences des variables de sorties en fonction du pH (ANOVA)

La réalisation de l'ANOVA (figure 6) permet de mettre en évidence les variables qui ont des moyennes différentes pour certains niveaux de pH grâce à la significativité des p-value. Le code utilisé est alors :

```
for(i in 3:length(Laqueuille_ext)){  
  res.aov <- aov(Laqueuille_ext[,i] ~ pH, data = Laqueuille_ext)  
  print(summary(res.aov)) # resume de l'analyse  
}
```

Figure 6 : Script Réalisation ANOVA (site de Laqueuille)

L'exécution de ce script sur les trois sites permet de déterminer quelles variables est significativement différentes en fonction du pH :

Tableau 1 : p-value par site et variable, quand celles-ci sont significatives

Groupe	Variable	Laqueuille pâturage	Theix fauche NPK	Theix fauche sans fertilisation
Sol C	wcorganic	< 2e-16	0.0272	/
Sol Ammonium	namm	< 2e-16	< 2e-16	< 2e-16
Sol Nitrate	nnit	< 2e-16	1.7e-10	/
Emissions N ₂ O	fn2osoilatm	2e-08	4.05e-12	/
Emissions NH ₃	fnh3_soil	< 2e-16	< 2e-16	< 2e-16
Lixivation N	leachsum	1.14e-07	< 2e-16	/
Sol N organique	wnorganic	< 2e-16	< 2e-16	/

Afin de savoir pour quels niveaux de pH cette différence est significative, la comparaison des moyennes 2 à 2 est effectuée en représentant la méthode de Tukey (figure 7) et en exécutant un t-test (figure 8).

```
plot(TukeyHSD(res.aov))
```

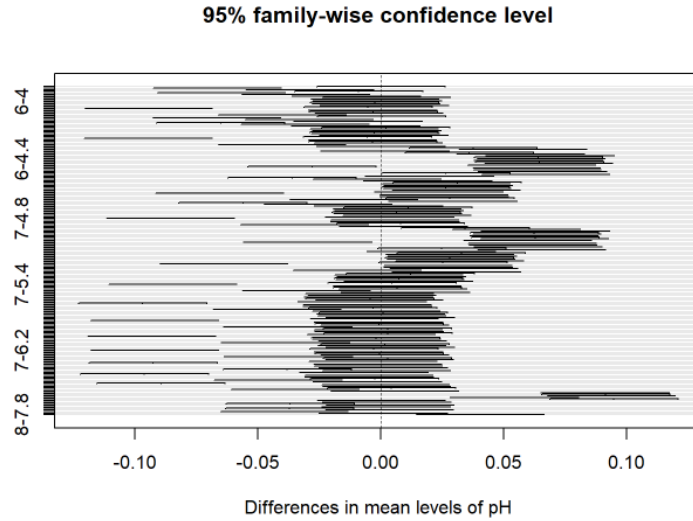


Figure 7 : Résultat méthode de Tukey avec la variable worganic et le site de Laqueuille

```
pairwise.t.test(laqueuille_ext$worganic, Laqueuille_ext$pH, p.adjust="bonf")
```

```
##
## Pairwise comparisons using t tests with pooled SD
##
## data: Laqueuille_ext$worganic and Laqueuille_ext$pH
##
##      4      4.2    4.4    4.6    4.8    5      5.2    5.4    5.6    5.8    6      6.2    6.4    6.6    6.8    7      7.2    7.4    7.6    7.8
## 4.2 1.00000 -      -      -      -      -      -      -      -      -      -      -      -      -      -      -      -      -      -
## 4.4 < 2e-16 < 2e-16 -      -      -      -      -      -      -      -      -      -      -      -      -      -      -      -
## 4.6 0.01579 0.01400 5.2e-05 -      -      -      -      -      -      -      -      -      -      -      -      -      -      -
## 4.8 1.00000 1.00000 5.9e-13 1.00000 -      -      -      -      -      -      -      -      -      -      -      -      -      -
## 5 < 2e-16 < 2e-16 1.00000 0.00018 3.8e-12 -      -      -      -      -      -      -      -      -      -      -      -      -
## 5.2 0.00621 0.00547 0.00016 1.00000 0.64955 0.00053 -      -      -      -      -      -      -      -      -      -      -      -      -
## 5.4 1.00000 1.00000 1.5e-12 1.00000 1.00000 9.7e-12 0.05474 -      -      -      -      -      -      -      -      -      -      -      -
## 5.6 1.00000 1.00000 < 2e-16 0.00403 1.00000 < 2e-16 0.00148 1.00000 -      -      -      -      -      -      -      -      -      -      -
## 5.8 1.00000 1.00000 3.2e-16 0.05925 1.00000 2.5e-15 0.02493 1.00000 1.00000 -      -      -      -      -      -      -      -      -      -
## 6 1.00000 1.00000 < 2e-16 0.03321 1.00000 7.0e-16 0.01356 1.00000 1.00000 1.00000 -      -      -      -      -      -      -      -      -
## 6.2 1.00000 1.00000 3.3e-16 0.06061 1.00000 2.6e-15 0.02553 1.00000 1.00000 1.00000 1.00000 -      -      -      -      -      -      -      -
## 6.4 1.00000 1.00000 < 2e-16 0.04696 1.00000 1.5e-15 0.01951 1.00000 1.00000 1.00000 1.00000 1.00000 -      -      -      -      -      -      -
## 6.6 1.00000 1.00000 < 2e-16 0.00602 1.00000 < 2e-16 0.00226 1.00000 1.00000 1.00000 1.00000 1.00000 1.00000 -      -      -      -      -      -
## 6.8 1.00000 1.00000 8.6e-15 0.24908 1.00000 6.3e-14 0.10961 1.00000 1.00000 1.00000 1.00000 1.00000 1.00000 1.00000 -      -      -      -      -
## 7 < 2e-16 < 2e-16 0.02816 < 2e-16 < 2e-16 0.01054 4.1e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 -      -      -      -
## 7.2 1.00000 1.00000 8.5e-16 0.09076 1.00000 6.6e-15 0.03908 1.00000 1.00000 1.00000 1.00000 1.00000 1.00000 1.00000 1.00000 < 2e-16 -      -      -
## 7.4 1.00000 1.00000 6.3e-16 0.07981 1.00000 4.9e-15 0.03412 1.00000 1.00000 1.00000 1.00000 1.00000 1.00000 1.00000 1.00000 < 2e-16 1.00000 -
## 7.6 1.00000 1.00000 < 2e-16 0.02430 1.00000 3.5e-15 0.00976 1.00000 1.00000 1.00000 1.00000 1.00000 1.00000 1.00000 1.00000 < 2e-16 1.00000 -
## 7.8 0.1e-05 7.8e-06 0.05686 1.00000 0.00414 0.13853 1.00000 0.00719 1.5e-06 5.5e-05 2.5e-05 5.7e-05 4.0e-05 2.5e-06 0.00039 1.8e-11 1.0e-04 8.3e-05 1.6e-05 -
## 8 1.00000 1.00000 < 2e-16 0.01225 1.00000 < 2e-16 0.00475 1.00000 1.00000 1.00000 1.00000 1.00000 1.00000 1.00000 < 2e-16 1.00000 1.00000 1.00000 6.5e-06
```

Figure 8 : Résultat t-test avec la variable worganic et le site de Laqueuille

A l'aide de ces deux tests, il est possible de produire ce tableau qui explicite les différences de moyennes lorsque celles-ci sont significatives au seuil de 5% :

Tableau 2 : Significativité des différences de moyenne au seuil de 5% suite aux méthodes de Tukey et t-test

Groupe	Variabes	Laqueuille pâturage	Theix fauche NPK	Theix fauche sans fertilisation
Sol C	worganic	/	/	/
Sol Ammonium	namm	Si pH fort	Si pH très fort	Si pH fort
Sol Nitrate	nmit	Si pH fort	Si pH très fort	/
Emissions N ₂ O	fn2osoilatm	Si pH fort	Si pH très fort	/
Emissions NH ₃	fnh3_soil	Si pH fort	Si pH fort	Si pH fort
Lixivation N	leachsum	/	Si pH fort	/
Sol N organique	wnorganic	Si pH fort	Si pH fort	/

➤ **Comparaison des résultats avec la littérature scientifique (analyse graphique et AFC)**

Analyse graphique

La réalisation de l'analyse graphique (figure 9) permet de visualiser les tendances de l'évolution des variables étudiées au cours du temps et en fonction des différentes valeurs prises par le pH. Les résultats graphiques sont présentés dans les annexes 5 à 7.

```
#install.packages("dichromat")
library(dichromat)
tabColor <- topo.colors(18)

moyenne_nnit = matrix(nrow = nbSimulation, ncol = nbYear)
for (i in 0:nbSimulation-1){
  for (j in 1:nbYear){
    moyenne_nnit[i+1,j] <- mean(Laqueuille_ext$nnit[year == j & pH == i*0.2 + 4])
  }
}

rownames(moyenne_nnit) = seq(4, 8, 0.2)
trans_moyenne_nnit = t(moyenne_nnit)

boxplot(trans_moyenne_nnit, xlab = "Indice de pH", ylab = "Nitrate présent dans les sols en kg N m-2", main = "Bloxplot nni
t")
legendColor <- c()
for(j in 1:nbYear){
  points(moyenne_nnit[,j], type='l', lwd =3, lty=2, col=tabColor[j])
  legendColor <- c(legendColor, paste0("Year ", j))
}

legend(x = "bottomright", legend = legendColor, fill = tabColor, cex = 0.6)
}
```

Figure 9 : Script Réalisation analyse graphique (variable nnit - site de Laqueuille)

Variable fnh3_nsoil – Volatilisation de l'ammoniac

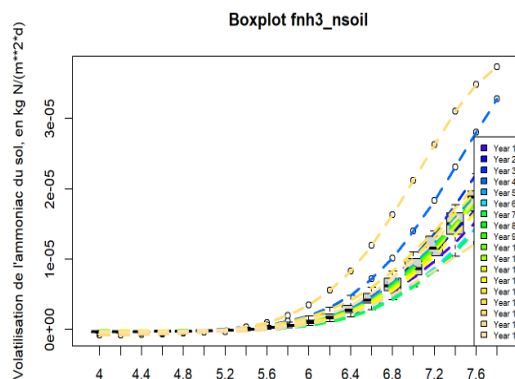


Figure 10 : Volatilisation de l'ammoniac en fonction du pH sur Laqueuille

Sur toutes les parcelles étudiées, les courbes qui représentent, pour une année donnée, les variations en fonction du pH, ont toutes le même sens de variations. Les valeurs n'étant toutefois pas les mêmes d'une année sur l'autre, il est possible de poser comme hypothèse qu'un des principaux facteurs explicatifs à la valeur des flux d'ammoniac (NH₃) est les conditions météorologiques de l'année. Il est de plus possible de constater que l'ensemble des courbes à tendance à croître au fur et à mesure que le

pH augmente. Cette croissance est bien plus conséquente lors de l'utilisation de fertilisants ou de présence d'animaux sur la parcelle (apports de pissats et de fèces) comme le montre la figure 10. Ainsi, dans l'étude de la parcelle de Theix sans fertilisation, la volatilisation de l'ammoniac, bien qu'augmentant avec la valeur du pH, reste toutefois très faible (de l'ordre de $10e-7$). Il est alors possible d'en conclure que d'une manière générale, plus le pH augmente, plus la volatilisation de l'ammoniac est importante mais que cela n'est réellement sensible qu'en situation où la parcelle est fertilisée ou avec pâturage. Cela est cohérent avec ce qu'il est possible de trouver dans la littérature scientifique puisque lorsque le pH est en dessous de 6.6, il est habituellement constaté qu'il y a très peu d'ammoniac disponible dans le sol et donc que sa volatilisation est quasi-nulle. Cependant avec les simulations PaSim, ce palier semble exister autour d'une valeur de pH de 6.

Variable nnit - Quantité de nitrate présent dans le sol

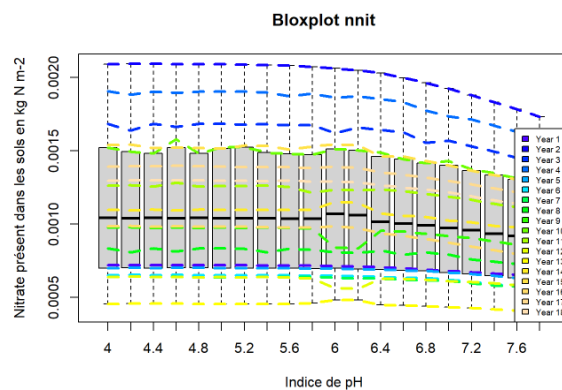


Figure 11 : Quantité de nitrate dans le sol en fonction du pH sur Laqueuille

A l'inverse de la volatilisation de l'ammoniac, il est ici possible de voir que la quantité de nitrate présent dans le sol décroît avec l'augmentation du pH, et ce indépendamment des années et donc des conditions météorologiques étudiées (figure 11). Cela est cohérent avec ce qu'il est possible de trouver dans la littérature scientifique car un pH fort traduit une activité plus importante des plantes donc une accumulation de nitrate dans le sol moindre.

Variable fn2osoilatm – Emission de N₂O

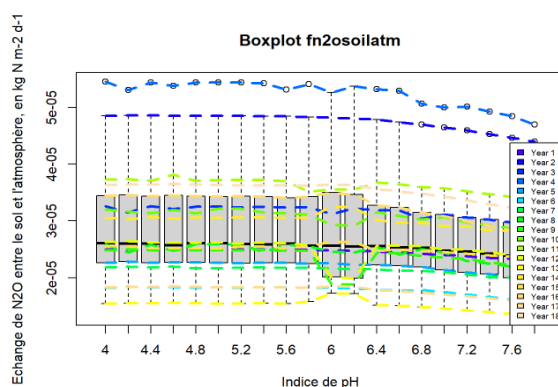


Figure 12 : Echange de N₂O en fonction du pH sur Laqueuille

Pour la parcelle de Laqueuille, autour d'un pH de 6, il est possible de constater une baisse du N₂O émis mais cela uniquement sur certaines années (figure 12). Ce phénomène est aussi observable sur les simulations Theix NPK lorsque le pH dépasse la valeur de 6.4. Il semblerait donc que les émissions de N₂O quotidiennes soient extrêmement corrélées avec les conditions météorologiques, mais que cette corrélation est moindre lorsqu'il y a fertilisation/pâturage sur un sol avec un pH avoisinant la neutralité. De plus, lorsqu'il y a fertilisation des sols ou pâturage, les valeurs ont tendance à décroître au fur et à mesure que le pH augmente, ce qui n'est pas le cas dans l'étude du site de Theix sans fertilisation.

Variable wnorganic – Quantité de nitrogène organique présent dans le sol

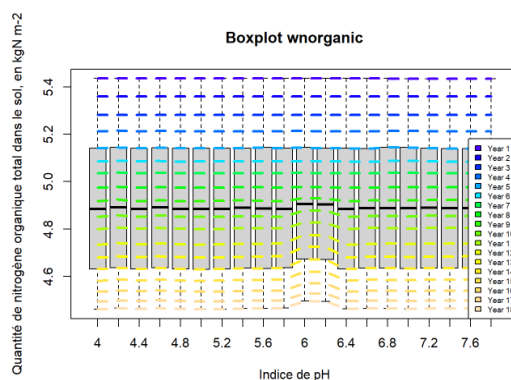


Figure 13 : Quantité de N organique dans le sol en fonction du pH sur Laqueuille

Sur les simulations des sites de Laqueuille et de Theix sans fertilisation, la quantité d'azote organique est quasi constante pour les différents pH (figure 13). Sans fertilisation ou avec pâturage, la quantité de nitrogène organique présent dans le sol dépend donc de l'année étudiée mais très peu du niveau de pH. De plus, dans ces deux cas, il est facilement visible que la quantité d'azote organique diminue d'année en année, ce qui peut s'expliquer facilement par un bilan négatif en azote (l'azote sort dans la biomasse fauchée ou ingéré par les animaux et aucun azote n'est apporté via des intrants). A l'inverse, dans le cas du site de Theix avec fertilisation NPK, les quantités de nitrogène présent dans le

sol sont initialement plus faibles mais augmentent d'année en année du fait d'un apport suffisant en fertilisant qui permet d'avoir un bilan d'azote positif.

AFC

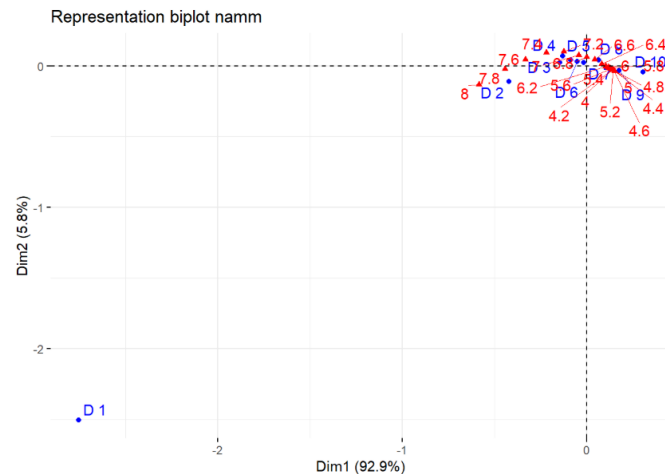


Figure 14 : Biplot AFC de la variable Ammonium (namm) sur le site de Laqueuille

La réalisation de l'AFC permet de connaître les tendances de l'évolution des variables en fonction des différentes valeurs prises par le pH. Les résultats graphiques sont présentés en annexe 8 à 10.

Lors de l'interprétation des résultats, il ne faut pas s'intéresser à la position absolue des variables dans le graphique mais plutôt à leur position relative (en particulier chercher à savoir de quel côté de l'axe des ordonnées elles se situent). Ainsi, la réalisation d'une AFC sur les 3 sites aux méthodes de fertilisation différentes a permis de mettre en évidence que :

- Sur l'ensemble des sites, les valeurs de concentration d'ammonium (namm, kg N/m²) élevées sont associées aux valeurs de pH faibles. En règle générale, la quantité d'ammoniac dans le sol a tendance à diminuer lorsque le pH devient de plus en plus fort (figure 14).
- Sur l'ensemble des sites, les valeurs d'ammoniac (fnh3_nsoil) élevées sont associées aux valeurs de pH fort. Cela implique que plus le pH d'un sol est alcalin, plus la volatilisation de l'ammoniac est importante, ce qui semble cohérent avec la conclusion.
- Sur le site de Laqueuille (pâturage extensif), les valeurs de concentration de nitrate (nmit) élevées sont étroitement liées avec les valeurs de pH faibles. La quantité de nitrate dans le sol est donc plus importante lorsque les sols sont acides.
- Sur les sites de Laqueuille (pâturage extensif) et de Theix (absence de fertilisation), les valeurs d'azote organique du sol (wnorganic) élevées sont associées aux valeurs de pH faibles. Ainsi la quantité d'azote organique présent dans le sol est moindre dans un sol plus alcalin lorsque la fertilisation est peu intense.
- Enfin, sur le site de Theix avec fertilisation NPK, les valeurs d'émissions de N₂O (fn2osoilatm) fortes, semblent pouvoir être associées aux valeurs de pH faibles. Dans une prairie fertilisée, les émissions de N₂O sont donc plus conséquentes lorsque le sol est acide.

➤ Etude sur l'effet du pH sur le rendement (Régression)

La réalisation d'une régression du pH sur le rendement par année a pour but de voir si l'effet du pH sur le rendement est modifié au regard de l'année de simulation. Dans un premier temps, les données de rendement simulées ont été représentées graphiquement comme lors de l'étude précédente. Les graphiques obtenus sont les suivants (figure 15) :

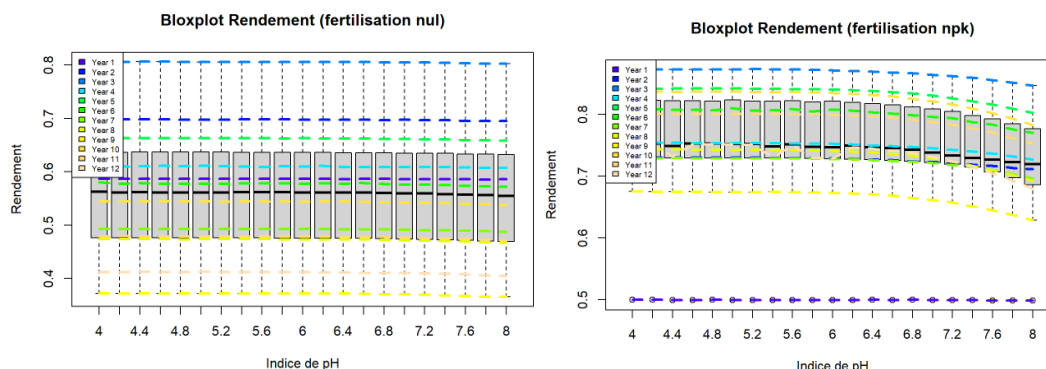


Figure 15 : Représentation du rendement annuelle (kg matière sèche/m²) en fonction du pH et par année, fertilisation NUL et NPK

Sur ces graphiques le rendement est stable malgré une modification du niveau de pH du sol lorsque la fertilisation est nulle et il est décroît lorsque le pH augmente sur la situation avec fertilisation. Cela est contre-intuitif puisqu'il était attendu d'avoir une augmentation du rendement avec le pH (c'est un des objectifs du chaulage). Il a donc été décidé d'effectuer une régression par année entre le pH et le rendement afin de voir si cette dernière était tout le temps identique ou bien si un déséquilibre dans le bilan d'azote pouvait expliquer ce résultat. Les régressions de chaque année semblent suivre les mêmes tendances c'est-à-dire une très légère augmentation du rendement lorsque le pH augmente de 4 à 6 puis une baisse de celui-ci comme le montre les représentations suivantes de 2 années choisies aléatoirement (figures 16 et 17) :

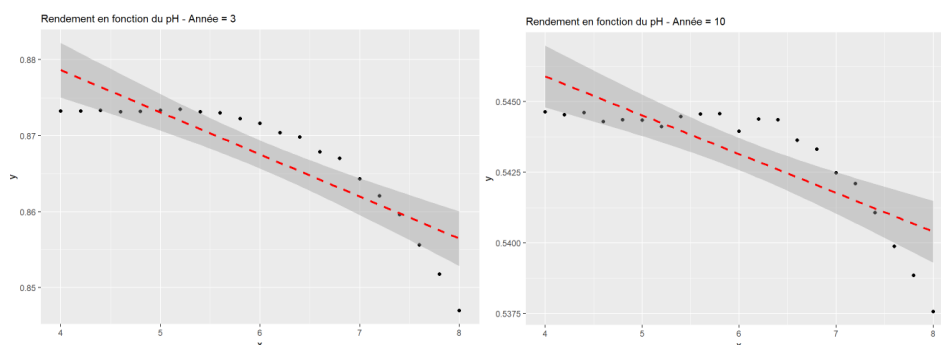


Figure 16 : Représentation de la régression linéaire entre le rendement et le pH pour les parcelles sans fertilisation

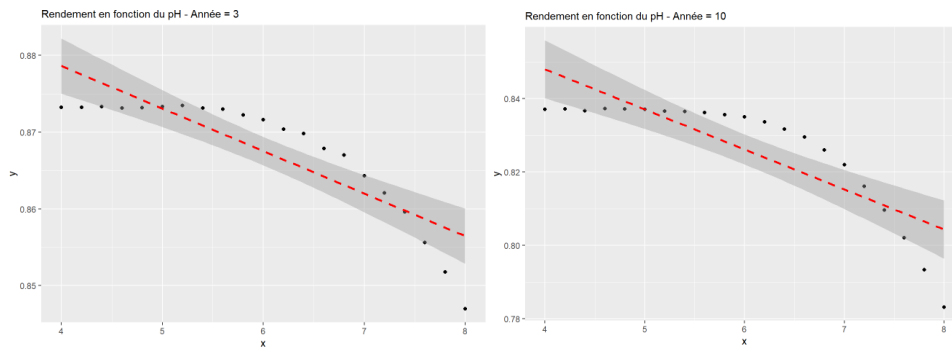


Figure 17 : Représentation de la régression linéaire entre le rendement et le pH pour les parcelles avec fertilisation NPK

➤ **Comparaison du rendement avec les données observées du terrain (t-test)**

Enfin, la réalisation d'un t-test permet de comparer les simulations des rendements avec les mesures du terrain. Tout d'abord, les valeurs simulées et observées sont représentées graphiquement au cours du temps (figure 18).

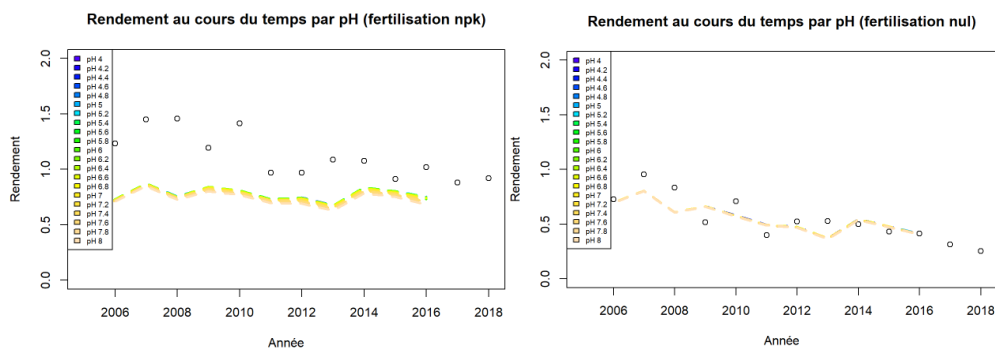


Figure 18 : Représentation du rendement simulé et observé au cours du temps en fonction du pH. Les valeurs simulées du rendement par année sont représentées par des droites colorées et les données réelles le sont par des points blancs.

Il apparaît que les différences de rendement des données simulées sont infimes en fonction des différents pH c'est pourquoi les droites représentant chaque valeur de pH se superposent. De plus, à première vue, l'estimation du rendement semble être plus proche de la réalité lorsqu'il n'y a pas de fertilisation, alors que les simulations NPK semblent sous-estimer le rendement.

Pour confirmer ou non ces hypothèses, des t-test sont mis en place afin de comparer les moyennes des différents échantillons de données. Les résultats renvoyés confirment les tendances dégagées par les graphiques : les moyennes des valeurs simulées sont statistiquement inférieures aux données réelles. Dans le cas de fertilisation NPK, les moyennes sont significativement différentes ($p < 0,001$) mais lorsqu'il y a absence de fertilisation ce n'est pas le cas ($p < 0,83$).

```
t.test(as.numeric(Theix_reel$NPK), y = Theix_npk$wshotsum, alternative="two.sided")

##
## Welch Two Sample t-test
##
## data: as.numeric(Theix_reel$NPK) and Theix_npk$wshotsum
## t = 6.4416, df = 12.248, p-value = 2.907e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.2487566 0.5021869
## sample estimates:
## mean of x mean of y
## 1.1210135 0.7455417

t.test(as.numeric(Theix_reel$NPK), y = Theix_npk$wshotsum, alternative="less")

##
## Welch Two Sample t-test
##
## data: as.numeric(Theix_reel$NPK) and Theix_npk$wshotsum
## t = 6.4416, df = 12.248, p-value = 1
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##   -Inf 0.4791837
## sample estimates:
## mean of x mean of y
## 1.1210135 0.7455417
```

Figure 19 : Résultats des Welch t-tests sur le rendement pour la parcelle avec fertilisation NPK

```
t.test(as.numeric(Theix_reel$NUL), Theix_nul$wshotsum, alternative="two.sided")

##
## Welch Two Sample t-test
##
## data: as.numeric(Theix_reel$NUL) and Theix_nul$wshotsum
## t = -0.21004, df = 12.423, p-value = 0.837
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1367718 0.1126376
## sample estimates:
## mean of x mean of y
## 0.5464922 0.5585593

t.test(as.numeric(Theix_reel$NUL), y = Theix_nul$wshotsum, alternative="less")

##
## Welch Two Sample t-test
##
## data: as.numeric(Theix_reel$NUL) and Theix_nul$wshotsum
## t = -0.21004, df = 12.423, p-value = 0.4185
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##   -Inf 0.09003796
## sample estimates:
## mean of x mean of y
## 0.5464922 0.5585593
```

Figure 20 : Résultats des t-tests sur le rendement pour la parcelle sans fertilisation

4. Discussion

Selon la littérature, les résultats attendus étaient les suivants (Y. Wang and al., 2017 et Catherine Hénault et al., 2019) :

- Au voisinage d'un pH neutre, la conversion microbienne de l'ammoniac (NH₃) en nitrate est rapide et donc sa volatilisation importante.

- Dans les sols acides, les émissions de N₂O augmentent. Cela explique que le chaulage des sols acides jusqu'à la neutralité du pH rende la réduction du N₂O plus efficace et diminue les émissions de N₂O du sol.
- Lorsque le pH est inférieur à 6.5, l'ammoniac est peu présent dans le sol et il y a donc peu de perte par volatilisation. Lorsque le pH est supérieur à cette valeur, la concentration en NH₃ augmente ainsi que, par conséquent, le risque de volatilisation.
- Plus le pH est fort meilleur est le rendement

Nombreuses de ces hypothèses se sont avérées être confirmées lors de cette étude :

- Le site de Laqueuille, prairie pâturée avec un faible chargement et pas de fertilisation, a des valeurs de nitrate (n_{nit}) élevées qui sont étroitement liées avec les valeurs de pH faibles. La quantité de nitrate dans le sol est donc plus importante lorsque les sols sont acides.
- Sur les sites de Laqueuille et de Theix sans fertilisation, les valeurs d'azote organique (w_{norganic}) élevées sont associées aux valeurs de pH faibles. Ainsi la quantité de nitrate organique présent dans le sol est moindre dans un sol plus alcalin lorsque la fertilisation est peu intense.
- Enfin, sur le site de Theix avec fertilisation NPK, les valeurs d'émissions N₂O (f_{n2o soil atm}) fortes semblent pouvoir être associées aux valeurs de pH faibles. Dans une prairie fertilisée, les émissions de N₂O sont donc plus conséquentes lorsque le sol est acide.

Les études menées permettent de conclure que dans les 3 situations, un certain nombre de variable fluctue en fonction des valeurs de pH (tableau 2). Dans la littérature scientifique, il est possible de trouver que lorsque le pH est en dessous de 6.6, il n'y a pas d'ammoniac dans le sol, ainsi, sa volatilisation est quasi nulle. Ce seuil se situe à 6 avec PaSim. En conséquence, plus le pH est fort, plus la volatilisation de l'ammoniac est importante et donc plus la quantité de nitrate présent dans le sol sera faible, ce qui implique une baisse de la quantité de nitrogène présent dans le sol. Ainsi, plus le pH est élevé, moins la quantité de nitrogène présent dans le sol devrait être importante. Les études de Theix avec fertilisation NPK et sans fertilisation sont en adéquation avec ce dernier point mais ce n'est pas le cas de l'étude de la quantité de nitrogène à Laqueuille puisque cette dernière reste constante, quel que soit le pH (elle varie uniquement en fonction des années). Enfin, pour toutes les simulations et comme dans la réalité, le modèle PaSim renvoie seulement une tendance que les sols acides émettent plus de N₂O.

En règle générale, la quantité d'ammoniac dans le sol a tendance à diminuer lorsque le pH devient de plus en plus fort. Cela implique que plus le pH d'un sol est alcalin, plus la volatilisation de l'ammoniac est importante, ce qui semble cohérent avec la conclusion précédente.

Enfin l'étude a mis en évidence que bien que le rendement des simulations du modèle PaSim soit plus important lorsque l'étude porte sur une prairie avec fertilisation NPK qu'en l'absence de fertilisation, il semble peu varier en fonction des différents niveaux de pH ce qui va à l'encontre de ce qu'il est possible de trouver dans la littérature scientifique. Cette tendance est même contraire à ce qu'il était possible de prévoir, pour des pH faibles, le rendement semble être stable mais lors de l'étude du site de Theix avec fertilisation NPK, le rendement baisse contrairement aux attentes. De plus la comparaison de ces données à la réalité a mis en évidence que lorsqu'il n'y a pas de fertilisation, le modèle PaSim simule bien le rendement réel mais ce n'est pas le cas lors de l'étude d'une fertilisation NPK, le rendement étant alors minimisé.

Cela met en évidence l'une des limites du modèle PaSim, l'analyse de sensibilité permettra ainsi de savoir quelles variables d'entrée étudiée influe le plus sur le rendement. De plus si la durée du stage le permet il pourrait être intéressant de regarder les équations du modèle afin de savoir comment cela pourrait être amélioré dans le modèle.

III. Analyse de sensibilité du modèle PaSim

L'analyse de la sensibilité d'un modèle permet de quantifier l'influence d'une entrée ou d'un groupe d'entrée sur la sortie d'un code ou d'un système. Elle doit reposer sur un grand nombre de simulations afin d'estimer l'effet moyen de la variation d'un paramètre sur la sortie observée.

Certaines étapes semblent essentielles à toute procédure car elles sont présentes dans la majorité d'entre elles :

- Quantifier l'incertitude dans chaque entrée
- Identifier la ou les sorties à analyser
- Exécuter le modèle un grand nombre de fois
- Calculer une mesure de sensibilité qui dépendra de la méthode utilisée

1. Lancement des simulations

Cette étude porte sur la sensibilité du modèle à une modification de pH. Afin de connaître l'influence de cette variable sur les sorties il est nécessaire de sélectionner d'autres variables d'entrées liées au sol, de les faire également varier dans les différentes simulations, dans le but de comparer l'effet du pH sur les sorties du modèle par rapport à celui des variables choisies. Cette sélection a été effectuée grâce à une étude théorique du modèle et à une analyse graphique de différents paramètres d'entrée, ceux qui ont été ainsi sélectionnés sont les suivants :

- Le pH en 9 niveaux (4 à 8 avec un pas de 0.5)
- Le stock de carbone organique (c'est-à-dire le SOC) en 3 niveaux (faible, moyen et fort)
- La proportion de légumineuse en 5 niveaux (0 à 20% avec un pas de 5% de légumineuse dans le couvert végétale)
- La texture du sol, en particulier la proportion de sable dans le sol en 9 niveaux (10 à 90% avec un pas de 10% de sable dans le sol)
- La profondeur de la première couche du sol en 4 niveaux (20 à 50 cm avec un pas de 10 cm)
- La quantité de précipitation journalière sera également considérée, elle est à la fois un paramètre d'entrée et une variable de sortie. Ainsi, elle pourra être considéré comme une variable témoin.

De la même manière, il a fallu s'intéresser aux variables clés impliquées dans le cycle de l'azote, du pH, du rendement et des émissions afin de déterminer quelles sorties extraire. Celles qui ont été sélectionnées sont les suivantes :

- L'année, **year**
- Les précipitations, **pa**, en mm/j
- La production primaire brute, **gpp**, en kgC/m², (il s'agit d'une somme)
- La quantité de carbone totale dans la plante, **Cplant** (substrat + structure)
- La quantité d'azote totale dans la plante, **Nplant** (substrat + structure)
- Le rendement, **wshtotsum**, en kg (de matière sèche)/m², (il s'agit d'une somme)
- La quantité d'azote issue des fauches, **yieldnsum**
- La fixation de N biologique, **bnfsum**
- La minéralisation du sol en azote, **fmintot**
- Le stock de carbone organique dans le sol, **wcorganic**, en kg (de carbone)/m²
- La quantité de N₂O issu de la dénitrification, **nn2odenit**

- Les flux de N₂O, **fn2osoilatm**
- Les émissions de N₂O, **n2oemissionsum**, en kg (d'azote)/m²
- La volatilisation du NH₃ du sol, **fnh3_nsoil**
- Le lessivage de l'azote du sol, **leachsum**
- Les pertes annuelles d'azote dues à la dénitrification, **n2denitrifsum**, en kg (d'azote)/m²
- La respiration de l'écosystème, **RECO**, en kg (de carbone)/m²
- La minéralisation journalière de l'azote du sol, **fmintot**, en kg (d'azote)/m²

➤ **Méthode de Monte Carlo**

D'après le schéma de décision (figure 21), il est judicieux de s'intéresser à la méthode de Monte Carlo car cette méthode est très efficace, peu coûteuse en temps de calcul et est une des rares qui permet de déterminer la sensibilité des sorties aux variables d'entrée.

Pour la réaliser, il est nécessaire de calculer des coefficients de corrélations entre les différentes variables. Sous R, il est possible de les calculer aisément grâce à la fonction `cor()` qui nécessite une matrice des sorties ainsi que le nom d'une méthode de calcul en entrée. Cette méthode renvoie alors la matrice de corrélation entre chaque variable de la table d'entrée. Avec cette fonction, la méthode de calcul des coefficients de corrélation peut être : *Kendall*, *Pearson* ou *Spearman*. La méthode de *Pearson* ne peut être effectuée que lorsque les données sont appariées c'est-à-dire lorsqu'il n'existe pas de lien direct entre les différentes observations, ce qui n'est pas le cas lorsque les données évoluent dans le temps puisque la valeur d'une variable à l'instant $t+1$ va dépendre de la valeur de cette même variable à l'instant t . Les méthodes de *Spearman* et *Kendall* renvoient des résultats tout aussi significatifs, cependant la méthode de *Kendall* est coûteuse en temps de calcul c'est pourquoi durant cette étude la méthode de Spearman sera utilisée pour calculer les coefficients de corrélation.

➤ **Représentation graphique**

Lors d'une analyse de sensibilité telle que celle de Monte Carlo, les indices en jeu peuvent être représentés sous la forme d'un diagramme en bâtons vertical (ou horizontal) sous R grâce à la fonction `barplot()`. Pour cela il est nécessaire que les indices soient en valeurs absolues et que la somme des indices d'une variable précise soit égale à 1. Ainsi, la représentation graphique permettra de mettre en évidence le pourcentage de contribution de chaque variable d'entrée sur les variables de sortie et/ou inversement.

➤ **Méthode de Saltelli**

Les recherches effectuées ont mis en évidence qu'en règle générale, les analyses de sensibilité se faisaient sur un vecteur d'entrées et un vecteur de sortie. Or dans le cadre de ce stage, les valeurs prises par les variables évoluent au cours du temps et donc une même variable ne va pas prendre une seule valeur mais une valeur par pas de temps. Afin de prendre en compte cela il sera donc nécessaire d'utiliser des matrices à l'instar de vecteurs. Une analyse de sensibilité entre une variable d'entrée et les variables de sorties sélectionnées avec un coefficient par variable et par jour pourra alors être résumé par une matrice de taille 14×4380 chaque ligne représentant une variable de sortie (à l'exception des sorties biomasse (`wshtotsum`) et rendement (`yieldnsum`) pour lesquelles l'étude des valeurs par jour n'est pas pertinente) et chaque colonne une journée de simulation. Cette matrice pourra alors être représentée grâce à une fonction réalisée sous R présentées dans les figures 22 et 23. Cette représentation pourra également être effectuée par année pour les deux variables laissées de côté en considérant leurs valeurs prises au dernier jour de l'année et la moyenne des autres variables de sortie.

```

# RepresentationCoefficientsSaltelli permet de représenter l'importance de différentes variables à l'aide d'un tableau de coefficients
# data = une matrice d'entrée contenant les coefficients qu'on souhaite représenter chaque colonne représentant une journée
# (dans notre cas c est la matrice de tous les coefficients de corrélation entre les variables de sortie et une variable d'entrée au cours du temps
# ou au contraire celle de tous les coefficients entre les variables d'entrée sur une sortie au cours du temps)
# row = la liste des variables qu'on souhaite analyser au cours du temps

RepresentationCoefficientsSaltelli <- fonction(data,row){
  data[is.na(data)] <- 0
  cor <- abs(data)
  nbcol <- ncol(cor)
  colnames(cor) <- c(1:nbcol)

  # transformation des données pour que la somme de chaque colonne soit égale à 1
  for (i in 1:nbcol){
    somme <- sum(cor[,i])
    if(somme == 0){
      somme <- 1
    } #dans le cas où toutes les valeurs d'une ligne sont nulles on veut des 0 non des NA
    for (j in 1:length(row)){
      cor[j,i] <- cor[j,i]/somme
    }
  }

  # la table des corrélations devient une table de corrélation cumulée
  for(k in 2:length(row)){
    cor[k,] <- cor[k,] + cor[k-1,]
  }

  tabcolor <- terrain.colors(length(row))

  plot(x=cor[1,],type="l", ylab="corrélation des variables", xlab="Temps en jours", ylim=c(0,1), col=tabcolor[1])
  for (k in 2:length(row)){
    points(x=cor[k,],type="l", ylab="corrélation des variables", xlab="Temps en jours", ylim=c(0,1), col=tabcolor[k])
  }

  legend(x="topleft", legend=row,fil=tabcolor,cex=0.6)
}

```

Figure 22 : Script de la fonction de la représentation de la méthode Saltelli

```

RepresentationCoefficientsSaltelli_annee <- fonction(data,row,year){
  data[is.na(data)] <- 0
  cor <- abs(data)
  nbcol <- ncol(cor)
  colnames(cor) <- c(1:nbcol)

  # transformation des données pour que la somme de chaque colonne soit égale à 1
  for (i in 1:nbcol){
    somme <- sum(cor[,i])
    if(somme == 0){
      somme <- 1
    } #dans le cas où toutes les valeurs d'une ligne sont nulles on veut des 0 non des NA
    for (j in 1:length(row)){
      cor[j,i] <- cor[j,i]/somme
    }
  }

  # la table des corrélations devient une table de corrélation cumulée
  for(k in 2:length(row)){
    cor[k,] <- cor[k,] + cor[k-1,]
  }

  tabcolor <- terrain.colors(length(row))

  for(i in 1:year){
    debut <- (i-1)*365 + 1
    fin <- i*365 - 1
    duree <- c(debut:fin)
    titre <- paste0("Corrélation des variables, Année = ", i)
    plot(cor[1,duree], type="l", main = titre,ylab="Corrélation des variables", xlab="Temps en jours", ylim=c(0,1),col=tabcolor[1])
    for (k in 1:length(row)){
      points(x=cor[k,duree],type="l", col=tabcolor[k])
    }
    legend(x="topleft", legend=row,fil=tabcolor,cex=0.6)
  }
}

```

Figure 23 : Script de la fonction de la représentation de la méthode Saltelli par année

3. Application

➤ Méthode de Monte Carlo

En pratique, en calculant les corrélations de *Spearman* grâce à la fonction de R, les valeurs de la matrice de corrélation étaient un ensemble de NA. Cela s'explique par le fait que certains écart type étaient nuls et que le calcul par cette méthode nécessite de diviser par cette valeur. Ainsi, dans certains cas, une valeur NA apparaissait. Il a donc été nécessaire de calculer les coefficients de *Spearman* étape par étape. Tout d'abord les moyennes de toutes les variables de sortie autre que biomasse (wshtotsum) et rendement (yieldnsum) par année ont été calculées, les valeurs du dernier jour de l'année ont été sélectionnées pour les deux variables mises de côté. Ensuite, les valeurs de la table ont été transformées en rang (c'est le fonctionnement sur lequel repose la méthode de calcul des coefficients de corrélation

de Spearman) puis les écarts types de chaque variable de sortie ainsi que celui du pH ont été calculés tout comme la covariance entre les sorties et le pH grâce à la fonction cov() de R. Ainsi les coefficients de corrélation ont été calculés en divisant la covariance par la multiplication des écart type. Une fonction a été conçue pour cela (figure 24), elle effectue cette méthode en renvoyant les coefficients de corrélation d'une variable d'entrée ou de sortie sur l'ensemble des variables de sorties ou d'entrées étudiées.

```
# creation d une fonction pour calculer les coefficients de correlations de Spearman sans la fonction R (qui renvoie des valeurs NA quand certaines valeurs sont nulles)
# En entree nous avons l ensemble de nos simulations avec chacune de nos variables d entree et de sortie sur chaque colonne
# les valeurs prises par ces donnees chaque jours de l annee sur chaque ligne
# on souhaite obtenir une matrice de corrélation entre une liste de variables de sortie (ou d entree) et une variable d entree (ou de sortie)
# ces différentes corrélation devront etre calculées tous les jours afin de connaître l evolution des liens qui existent entre les différentes variables

# en entree nous avons data l ensemble de nos donnees, x l indice de la variable a analyser,
# y les indices des variables que l on souhaite représenter, nbyear le nombre d annee étudiée et nbsimulation le nombre de simu

Correlationspearman_jour <- fonction(data,x,y,nbyear,nbsimulation){
  nbDay <- nbyear*365
  cor <- matrix(nrow = length(y),ncol = nbDay)

  # on transforme les valeurs de la table en rang
  data_rank <- data
  for(j in 1:length(y)+1){
    data_rank[,j] <- rank(data[,j])
  }

  # calcul des coefficients de Pearson pour chaque annee
  nbcol <- nbday

  c <- c()
  r <- c()

  for (i in 1:nbcol){
    c <- cbind(c,r)
    r <- c()
    for(j in 1:length(y)){
      sigma_ph <- sd(data_rank[seq(i,nbday*nbsimulation,nbDay),x])
      sigma_sortie <- sd(data_rank[seq(i,nbday*nbsimulation,nbDay),y[j]])
      r <- c(r,cov(data_rank[seq(i,nbday*nbsimulation,nbDay),c(x,y[j])])[1,2]/sigma_ph/sigma_sortie)
    }
  }
  return(c)
}
```

Figure 24 : Script de la fonction renvoyant une matrice de corrélation de Spearman par jour

La matrice de corrélation (figure 25) est alors représentée sous la forme de diagrammes en bâtons afin d'avoir une idée visuelle de l'impact de chacune des variables d'entrée sur les sorties (figure 26) et la corrélation des sorties sur les entrées (figure 27). Cela permet de mettre en évidence que la précipitation est une bonne variable témoin puisque cette sortie est 100% corrélée avec l'entrée qui lui est associée. Les seules variables sur lesquelles le pH a un impact fort (supérieur à 10%) sont les variables nn2odenit [13.4%], fn2osoilatm [16.9%], n2oemissionsum [26.6%] et fnh3_nsoil [69.3%] soit les différents rejets d'émissions de N₂O et la volatilisation de l'ammoniac des sols. Ces tendances se confirment lors de l'analyse des figures 26 et 27. Cette étude met en évidence que certaines variables ont une corrélation infime avec le pH, c'est le cas de la minéralisation d'azote (fnmintot) [0.8%] et du carbone organique (wcorganic) [0.1%]. Enfin il est également possible de remarquer grâce à la figure 26 que la variable indiquant le pourcentage de légumineuses a un fort impact sur la variable bnfsum (fixation de l'azote atmosphérique par les légumineuses). De plus, il est possible de constater un fort effet des précipitations sur la minéralisation d'azote dans le sol (fnmintot) [18.3%], du N₂O issu de la dénitrification (nn2odenit) [12.2%] et de l'ammoniac (fnh3nsoil) [24.2%] soient la minéralisation du sol en azote, la quantité de N₂O issu de la dénitrification et la volatilisation du NH₃ du sol.

	pa	gpp	Nplant	Cplant	wshottsum	yieldnsum	bnfsum	fnmintot	wcorganic	nn2odenit	fn2osoilatm	n2oemissionsum	fnh3_nsoil	leachsum	n2denitrifsum	RECO
## soc	0.20276350	0.20669839	0.34195454	0.18113079	0.14684974	0.128224203	0.190187186	0.446321180	0.2006636	0.25291259	0.22310878	0.032727755	0.13071779	0.10053275	0.36079594	
## profondeur	0.29138038	0.35061584	0.31884769	0.31583306	0.33345532	0.006983042	0.345788153	0.451300452	0.2828195	0.04990251	0.03211691	0.022610474	0.2554427	0.14366859	0.40458658	
## texture	0.17853024	0.15075644	0.14538462	0.15996002	0.16005155	0.045767797	0.238756229	0.079867676	0.1439544	0.24607041	0.26016036	0.108551717	0.20586359	0.34057312	0.16652695	
## legumineuse	0.11021878	0.15296419	0.11925583	0.12973605	0.15027063	0.669713600	0.029572536	0.006740519	0.1011033	0.16861131	0.16130201	0.006810708	0.05654221	0.03055911	0.03906443	
## pH	0.03174918	0.04262315	0.03048670	0.03031184	0.04463784	0.020637138	0.008397771	0.001226345	0.1340477	0.25407543	0.26628112	0.692909696	0.05649810	0.04297395	0.01157103	
## pa	1	0.18535791	0.09634200	0.04407062	0.17502823	0.16473493	0.128674219	0.187298125	0.014543829	0.1374116	0.02842775	0.05703083	0.136381651	0.28583405	0.34169289	0.01045508

Figure 25 : Matrice de corrélation de Spearman

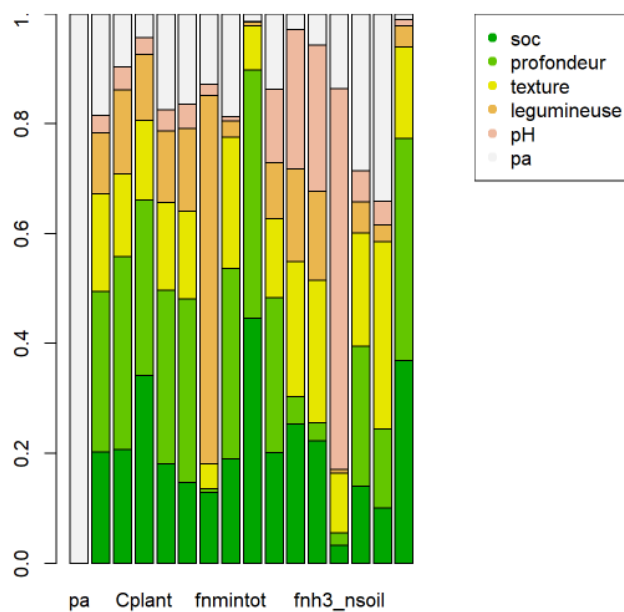


Figure 26 : Diagramme en bâtons de la contribution **des entrées sur les sorties** (axe des abscisses : pa, gpp, Nplant, Cplant, wshtotsum, yieldnsum, bnfsum, fnmintot, worganic, nn2odenit, fn2osoilatm, n2oemissionsum, fnh3_nsoil, leachsum, n2denitrifsum, RECO)

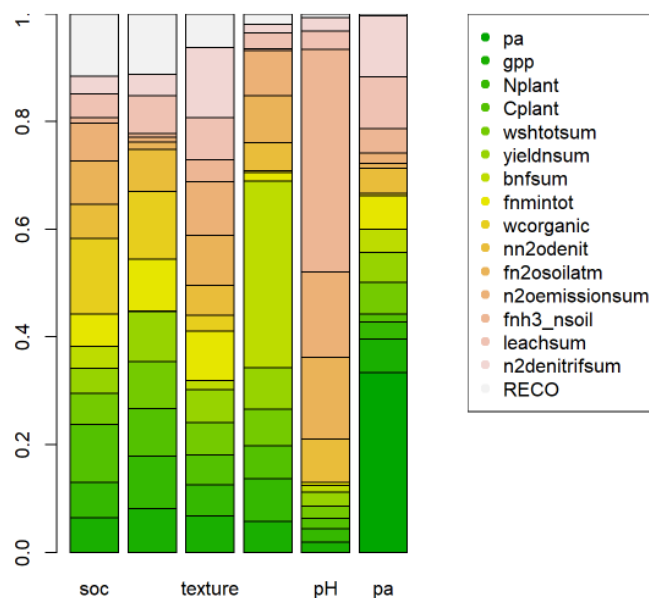


Figure 27 : Diagramme en bâtons de la contribution **des sorties sur les entrées** (axe des abscisses : soc, profondeur, texture, légumineuse, pH et pa)

➤ Méthode de Saltelli

Grâce à la représentation de Saltelli, les corrélations des variables peuvent être représentées au cours du temps comme le présente les figures des annexes 12 à 14. Dans cette partie, l'objectif est d'analyser l'impact du pH sur les sorties, ce graphique est représenté dans les figures 28.

Il est alors possible de remarquer que la variable de la volatilisation de l'ammoniac est celle qui est le plus impactée par la valeur du pH sur l'intégralité de l'étude. Cependant, au cours de chaque année, la corrélation entre les variables de sortie et le pH semble varier, cela semble en particulier du aux variations de la variable *fmintot*.

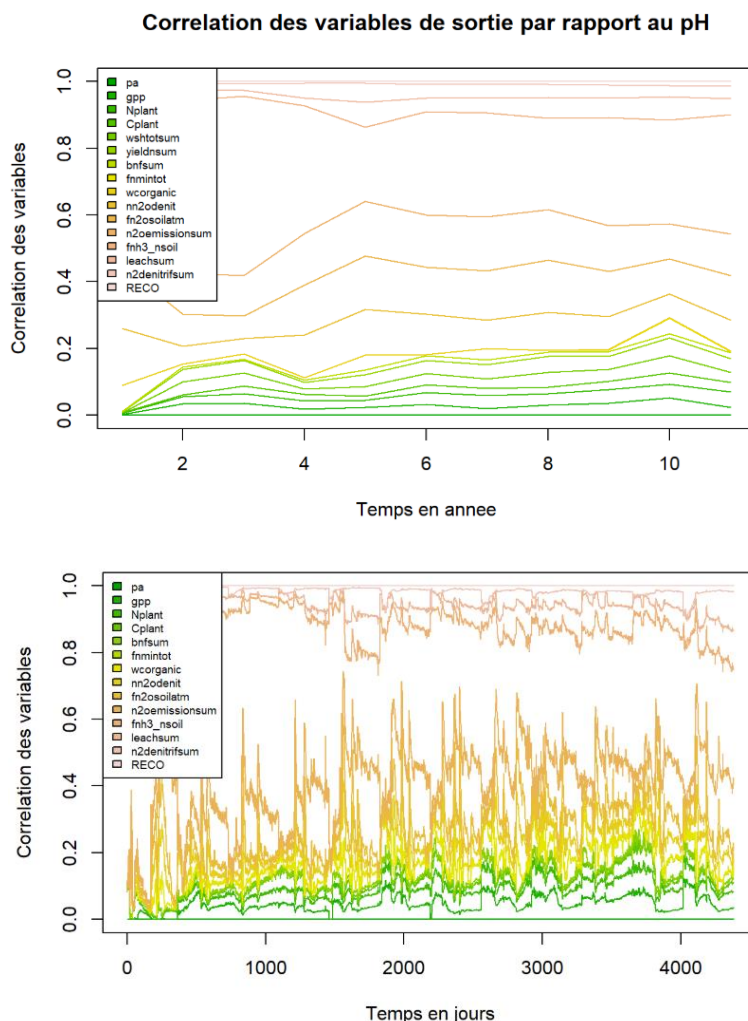


Figure 28 : Graphiques de Saltelli pour la corrélation des variables de sortie avec le pH au cours du temps (étude par année en haut puis par jour en bas)

De plus, la représentation de la corrélation des variables d'entrées en fonction de chaque variable de sortie (annexe 14) a pu être réalisée par jour pour la majorité des sorties étudiées à l'exception des variables *wshtotsum* et *yieldnsum* qui ont été représentées au cours des années, l'étude de l'évolution de ces variables par jour n'étant pas judicieux.

Ainsi, ces représentations ont permis de faire ressortir que la volatilisation de l'ammoniac est la variable la plus impactée par la valeur du pH sur l'intégralité de la simulation.

La représentation des variables *wshtotsum* et *yieldsum* renvoient les mêmes conclusions, il semblerait que le pH et les précipitations n'aient aucune influence sur les variables étudiées.

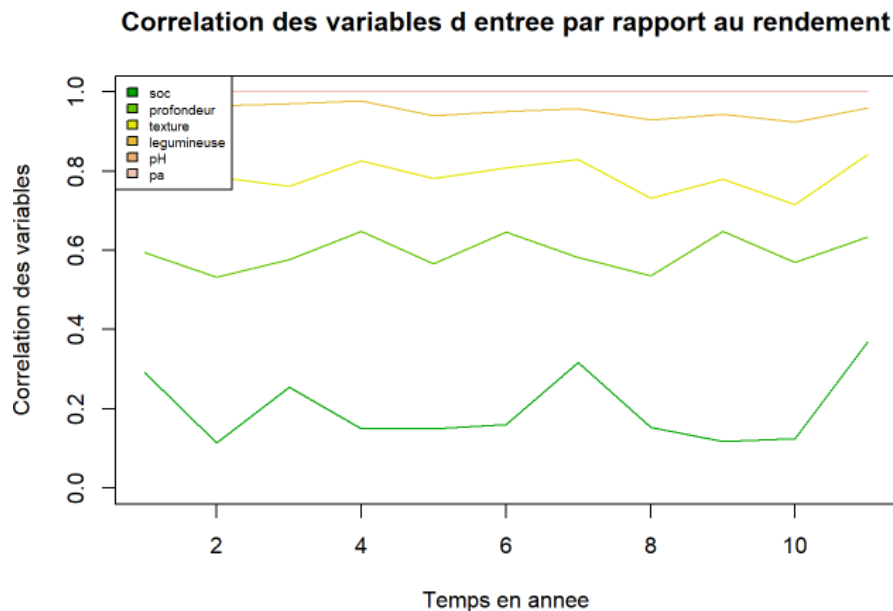


Figure 29 : Graphique de Saltelli de la corrélation des variables d'entrée avec le rendement au cours des années

4. Discussion

Ainsi, l'analyse de sensibilité avec la méthode de Monte Carlo a permis de mettre en évidence que le rendement n'est pas impacté par une modification du pH. Ce n'est cependant pas le cas des variables illustrant les rejets de N_2O et d'ammoniac dans l'air. Il semblerait que la variable représentant la volatilisation de l'ammoniac (fnnh3soil) soit de loin celle qui est la plus impactée par une modification de pH, suivi par les variables nn2odenit, fn2osoilatm et n2oemissionsum soient respectivement la quantité de N_2O issue de la dénitrification, les flux de N_2O et les émissions de N_2O .

Cela semble pouvoir être cohérent avec la littérature (Y. Wang and al., 2017 et Catherine Hénault et al., 2019) qui montre qu'avec l'acidité du sol, les émissions de N_2O augmentent. Les simulations confirment que le modèle pourrait de ce fait simuler un amendement de chaulage jusqu'à la neutralité du pH et ainsi reproduire les effets de chaulage sur les émissions de N_2O du sol.

En ce qui concerne l'effet du chaulage sur le rendement, il reste étonnant qu'à part les 4 variables des cycle gazeux d'azote, la corrélation entre le rendement et le pH soit si faible. Il serait attendu que lorsque le pH est fort, le rendement soit plus important. Et bien que les analyses de sensibilité montrent un léger effet du pH sur le rendement, l'impact des autres variables comme la précipitation et les conditions initiales du sol sont plus importantes.

Le modèle PaSim semble suite à cette analyse être un bon simulateur d'une modification de pH lorsque les émissions de GES sont étudiées mais ce n'est pas le cas lorsque l'étude porte sur le rendement.

De plus des variables telles que la quantité de carbone organique présent dans le sol (worganic) et la fixation d'azote (bnfsum) renvoient des tendances cohérentes avec ce qu'il pouvait être attendu : la première est extrêmement corrélée avec la variable soc, la seconde l'est avec la proportion de légumineuse. Cela permet de mettre en évidence que la modélisation par le modèle PaSim était pertinente.

Conclusion

Le modèle PaSim est représentatif de l'effet qu'a le pH sur les émissions de N₂O mais ce n'est pas le cas de son impact sur le rendement.

Bien que la réalisation de simulation avec le modèle PaSim ait permis de mettre en évidence que de nombreuses informations présentes dans la littérature scientifique soient prises en compte, il n'est pas possible de considérer que le modèle PaSim représente parfaitement l'impact qu'a une modification de pH sur le sol d'une prairie permanente. Bien que cela semble être le cas lors de l'étude d'une prairie fauchée sans fertilisation et sans pâturage, l'analyse de sensibilité effectuée sur la prairie de Theix avec fertilisation NPK met en évidence les limites de cette modélisation.

L'effet du pH est représentatif en ce qui concerne les flux de gaz azoté mais il n'a aucune conséquence sur le rendement ni sur de nombreuses autres variables qui devraient être impactées par une modification de pH. Cela peut être dû à une mauvaise représentation du pH dans les équations du modèle, il serait donc nécessaire de regarder celles portant sur le rendement et le pH plus en détail afin de perfectionner la prise en compte de l'impact du pH au sein du modèle PaSim. Ainsi, cela pourrait éventuellement permettre d'expliquer les écarts remarquables entre les valeurs observées et celles qui sont simulées.

Bibliographie

Agriculture et gaz à effet de serre : état des lieux et perspectives, 2010
Diane Vandaele, Amandine Lebreton, Benoît Faraco

Les effets du chaulage sur les prairies permanentes ou de longue durée, 2006
B. Fabre, F. Kockmann

Global Sensitivity Analysis. The Primer, 2008
Andrea Saltelli, Marco Ratto et al.

Management of soil pH promotes nitrous oxide reduction and thus mitigates soil emissions of this greenhouse gas, 2019
Catherine Hénault, Hocine Bourennane

La multifonctionnalité de la prairie pour la fourniture de services écosystémiques, 2012
B. Amiaud, P. Carrère

Nonlinear Sensitivity Analysis of Multiparameter Model Systems, 1978
R. I. Cukier, H. B. Levine et K. E. Shuler

Place et atouts des prairies permanentes en France et en Europe, 2012
J.-L. Peyraud, A. Peeters, A. De Vlieghe

Le protoxyde d'azote (N₂O), puissant gaz à effet de serre émis par les sols agricoles : méthodes d'inventaire et leviers de réduction, 2013
Amélie VIARD, Catherine Hénault et al.

Review of sensitivity analysis methods and experience Milestone, 2008
Anca Badea and Ricardo Bolado

Sensitivity and Uncertainty analysis, Theory, Volume I, 2003
Dan G. Cacuci

Soil pH as the chief modifier for regional nitrous oxide emissions: New evidence and implications for global estimates and mitigation, 2017
Yajing Wang, Jingheng Guo

Le stockage de carbone par les prairies, 2010
Armelle Gac, Jean-Baptiste Dollé et al.

Webographie

<http://www.sthda.com/>

<https://azote.info/environnement-et-azote/eviter-les-pertes-par-volatilisation.html>

<https://madamefinance.wordpress.com/2014/06/09/quadrant-ii-van-analyse-de-sensibilite/>

https://fr.wikipedia.org/wiki/Indice_de_Sobol

https://fr.wikipedia.org/wiki/Analyse_de_sensibilit%C3%A9

http://www.itse.be/statistique2010/co/234_Cours_ecart_moyen_median_absolu.html

<https://abcdr.thinkr.fr/comment-tester-la-conformite-a-une-loi-avec-le-test-de-kolmogorov-smirnov/>

<http://perso.ens-lyon.fr/lise.vaudor/test-de-wilcoxon-mann-whitney/>

https://rstudio-pubs-static.s3.amazonaws.com/133584_8726e36a0d4348d8918952505956e388.html

<https://www.futura-sciences.com/sciences/definitions/chimie-ph-222/>

Annexes

Annexe n°1 : Données initiales de l'exécution de PaSim (Site de Laqueuille)

β	Description
0.79	Latitude N (rad)
12.12	Time of the highest position of the sun (h.min)
0.000	Slope (rad)
0.000	Aspect (rad)
1040.0	Height above sea level(m)
2.0	Micromet reference height above soil surface (mm)
2.0	NH3 reference height above soil surface (m)
6	Number of soil layers (max=6)
20 100 200 400 600 900	Depth of soil layers (mm)
1000.0	Depth of lower soil boundary layer (mm).
0.60	Maximal canopy height (m)
5.00	Canopy height parameter
0.05	Clover fraction (kg/kg)
0.126 0.363 0.252 0.190 0.060 0.009	Relative root dry matter in different soil layers (-)
0.20	Main rooting depth (m)
0.54 0.54 0.68 0.69 0.7 0.8	Bulk density (kg/l)
0.19 0.19 0.23 0.25 0.25 0.37	Sand fraction of texture (-)
0.26 0.26 0.23 0.20 0.21 0.15	Clay fraction of texture (-)
0.55 0.55 0.54 0.55 0.54 0.48	Silt fraction of texture (-)
0.53 0.53 0.55 0.55 0.57 0.57	Saturated soil water content (m3/m3)
0.57	Saturated soil water content of lower soil boundary layer (m3/m3)
-385 -385 -374 -370 -379 -317	Air entry potential (mm)
-317	Air entry potential of lower soil boundary layer (mm)
8.7 8.7 8 7.5 7.6 6.2	Parameter b in psi (-)
6.2	Parameter b of lower soil boundary layer (-)
230. 230. 230. 205. 230. 230.	Saturated hydraulic conductivity (mm/d)
230	Saturated hydraulic conductivity of lower soil boundary layer (mm/d)
0.01	Parameter for the determination of the field capacity
0.33 0.33 0.33 0.33 0.33 0.33	Field capacity
0.218 0.218 0.226 0.204 0.173 0.173	Permanent wilting point
5.7	Soil pH
1.076 1.076 1.076 1.076 1.076 1.076	Parameter a for soil NH4+ partitioning
0.66 0.66 0.66 0.66 0.66 0.66	Parameter b for soil NH4+ partitioning

Annexe n°2 : Données initiales de l'exécution de PaSim (Site de Theix – fertilisation NPK)

./site_spe_block1_npk.par	Description
0.7980	Latitude N (rad)
12.12	Time of the highest position of the sun (h.min)
0.000	Slope (rad)
0.000	Aspect (rad)
882	Height above sea level(m)
2.0	Micromet reference height above soil surface (mm)
2.0	NH3 reference height above soil surface (m)
6	Number of soil layers (max=6)
20 100 200 300 400 600	Depth of soil layers (mm)
800	Depth of lower soil boundary layer (mm)
0.50	Maximal canopy height (m)
4.00	Canopy height parameter
0.016	Clover fraction (kg/kg)
0.118 0.152 0.198 0.25 0.206 0.076	Relative root dry matter in different soil layers (-)
0.5	Main rooting depth (m)
0.94 0.94 0.94 1.26 1.26 1.58	Bulk density (kg/l)
0.566 0.566 0.566 0.588 0.588 0.708	Sand fraction of texture (-)
0.194 0.194 0.194 0.168 0.168 0.098	Clay fraction of texture (-)
0.239 0.239 0.239 0.244 0.244 0.194	Silt fraction of texture (-)
0.566 0.566 0.57 0.475 0.475 0.375	Saturated soil water content (m3/m3)
0.375	Saturated soil water content of lower soil boundary layer (m3/m3)
-40.147 -40.147 -40.147 -125.352 -125.352 -170.775	Air entry potential (mm)
-170.775	Air entry potential of lower soil boundary layer (mm)
6.166 6.166 6.166 5.651 5.651 4.02	Parameter b in psi (-)
4.02	Parameter b of lower soil boundary layer (-)
500 500 500 433.289 433.289 310.77	Saturated hydraulic conductivity (mm/d)
310.77	Saturated hydraulic conductivity of lower soil boundary layer (mm/d)
0.01	Parameter for the determination of the field capacity
0.361 0.361 0.392 0.284 0.284 0.168	Field capacity
0.191 0.191 0.225 0.143 0.143 0.061	Permanent wilting point
6.096	Soil pH (kg/kg)
1.076 1.076 1.076 1.076 1.076 1.076	Parameter a for soil NH4+ partitioning
0.66 0.66 0.66 0.66 0.66 0.66	Parameter b for soil NH4+ partitioning

Annexe n°3 : Données initiales de l'exécution de PaSim (Site de Laqueuille – fertilisation nulle)

```

./site_spe_block1_nul.par Description
0.7980 Latitude N (rad)
12.12 Time of the highest position of the sun (h.min)
0.000 Slope (rad)
0.000 Aspect (rad)
882 Height above sea level(m)
2.0 Micromet reference height above soil surface (mm)
2.0 NH3 reference height above soil surface (m)
6 Number of soil layers (max=6)
20 100 200 300 400 600 Depth of soil layers (mm)
800 Depth of lower soil boundary layer (mm)
0.50 Maximal canopy height (m)
4.00 Canopy height parameter
0.151 Clover fraction (kg/kg)
0.118 0.152 0.198 0.25 0.206 0.076 Relative root dry matter in different soil layers (-)
0.5 Main rooting depth (m)
0.896 0.896 0.896 1.258 1.258 1.621 Bulk density (kg/l)
0.525 0.525 0.525 0.59 0.59 0.688 Sand fraction of texture (-)
0.19 0.19 0.19 0.156 0.156 0.092 Clay fraction of texture (-)
0.284 0.284 0.284 0.255 0.255 0.22 Silt fraction of texture (-)
0.58 0.58 0.584 0.475 0.475 0.364 Saturated soil water content (m3/m3)
0.364 Saturated soil water content of lower soil boundary layer (m3/m3)
-34.496 -34.496 -34.496 -122.299 -122.299 -188.737 Air entry potential (mm)
-188.737 Air entry potential of lower soil boundary layer (mm)
6.26 6.26 6.26 5.448 5.448 4.042 Parameter b in psi (-)
4.042 Parameter b of lower soil boundary layer (-)
500 500 500 422.349 422.349 266.489 Saturated hydraulic conductivity (mm/d)
266.489 Saturated hydraulic conductivity of lower soil boundary layer (mm/d)
0.01 Parameter for the determination of the field capacity
0.376 0.376 0.408 0.283 0.283 0.172 Field capacity
0.205 0.205 0.24 0.14 0.14 0.064 Permanent wilting point
6.107 Soil pH (kg/kg)
1.076 1.076 1.076 1.076 1.076 1.076 Parameter a for soil NH4+ partitioning
0.66 0.66 0.66 0.66 0.66 0.66 Parameter b for soil NH4+ partitioning

```

Annexe n°4 : Script représentation graphique du rendement

```

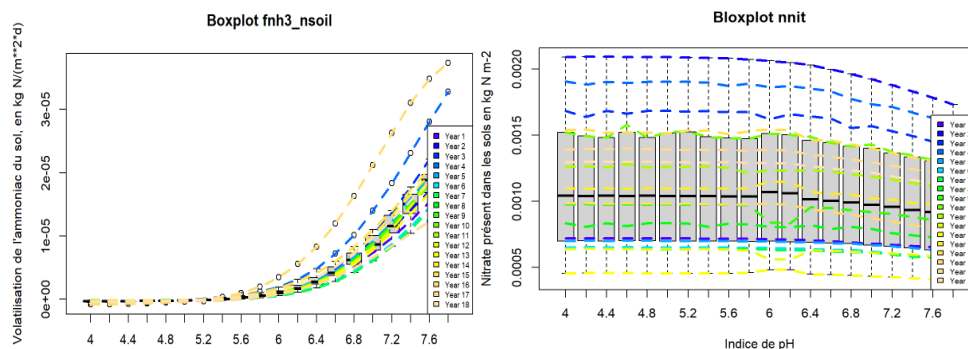
#install.packages("dichromat")
library(dichromat)
tabColor <- topo.colors(nbSimulation)

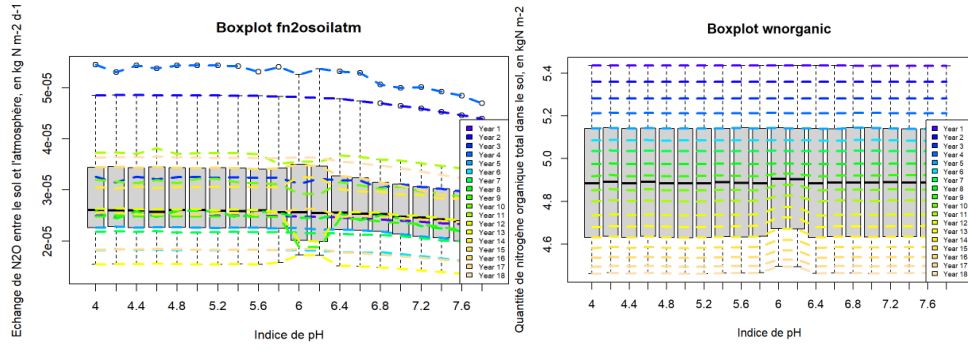
plot(Theix_reel$Annee, Theix_reel$NPK, type = "p", xlim = c(2005,2018), ylim = c(0,2), xlab = "Année", ylab = "Rendement", m
ain = "Rendement au cours du temps par pH (fertilisation npk)")
legendColor <- c()
for(i in 1:nbSimulation){
  j = (i-1)*0.2 + 4
  print(points(Theix_npk$year[Theix_npk$pH == j], Theix_npk$wshtotsum[Theix_npk$pH == j], type='l', lwd =3, lty=2, col=tabCo
lor[i]))
  legendColor <- c(legendColor, paste0("pH ", j))
}

legend(x = "topleft", legend = legendColor, fill = tabColor, cex = 0.6)

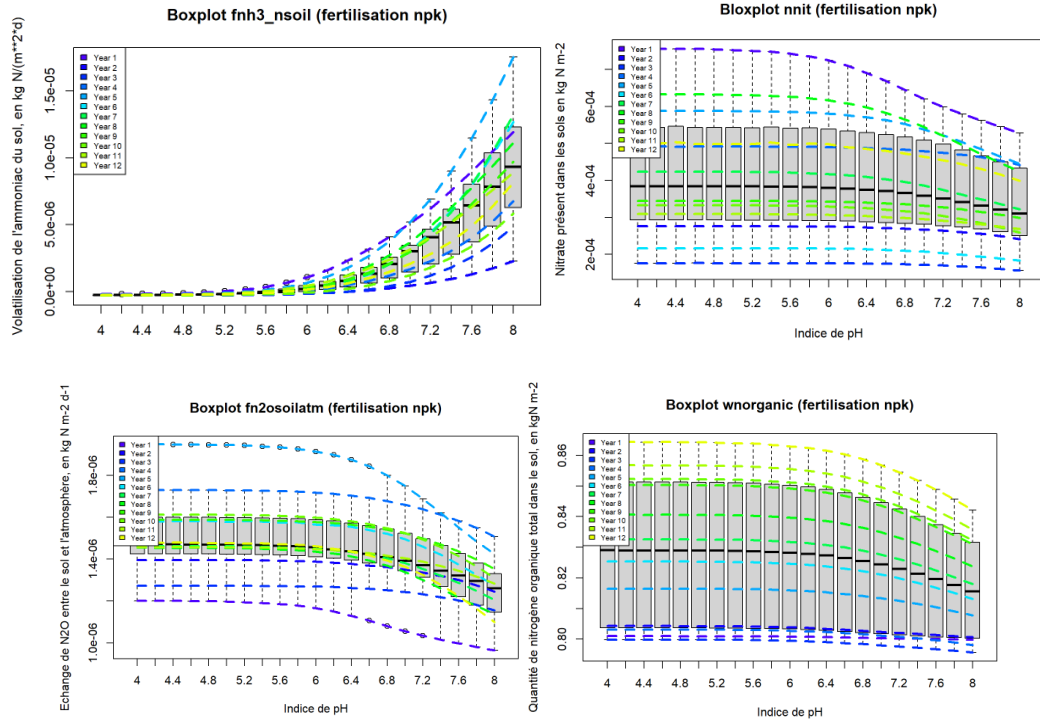
```

Annexe n°5 : Analyse graphique variable de Laqueuille

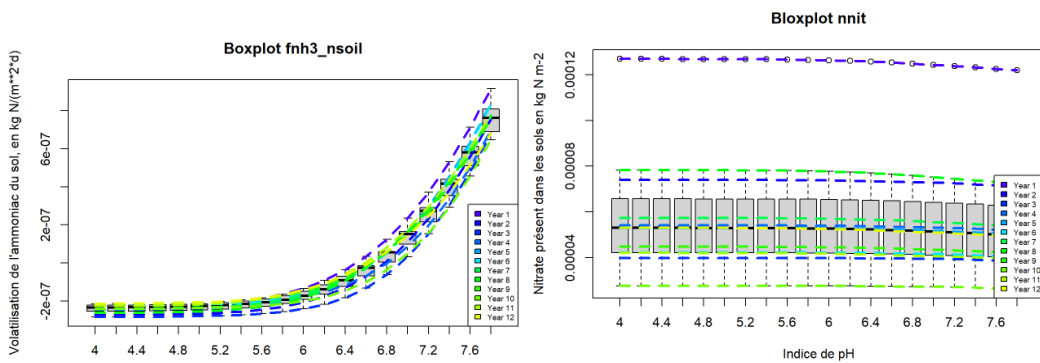




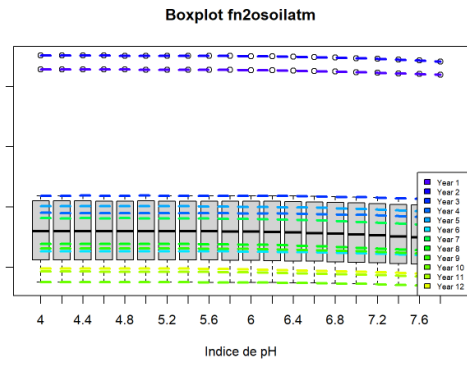
Annexe n°6 : Analyse graphique variable de Theix fertilisation NPK



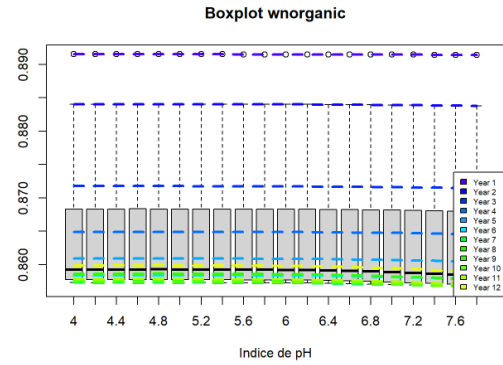
Annexe n°7 : Analyse graphique variable de Theix fertilisation NUL



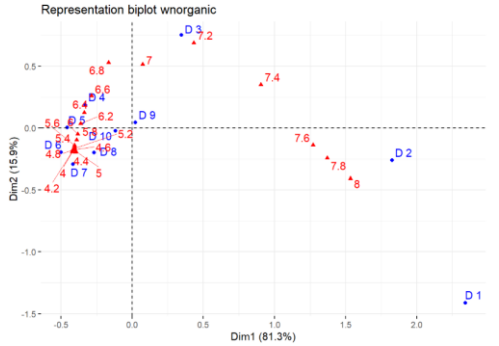
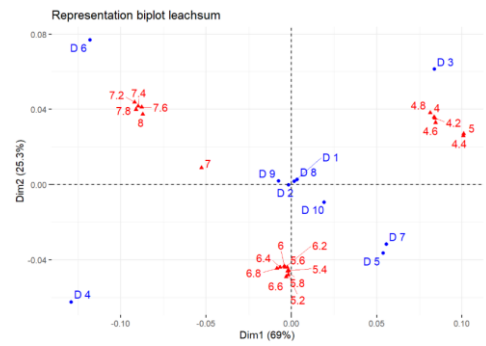
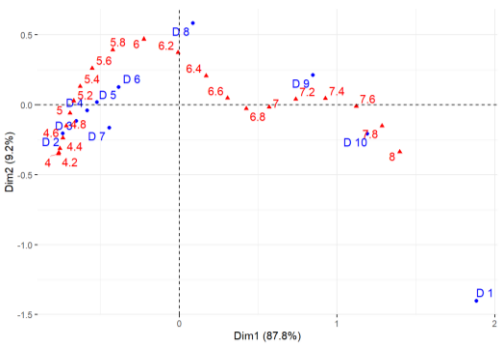
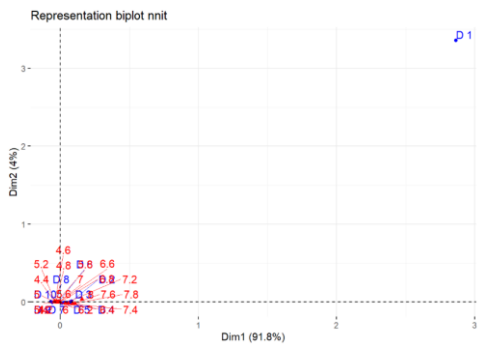
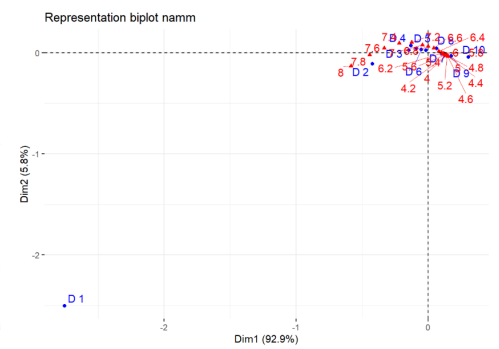
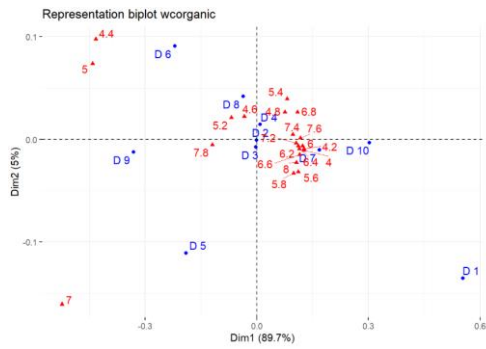
Echange de N2O entre le sol et l'atmosphère, en kg N m⁻² d⁻¹



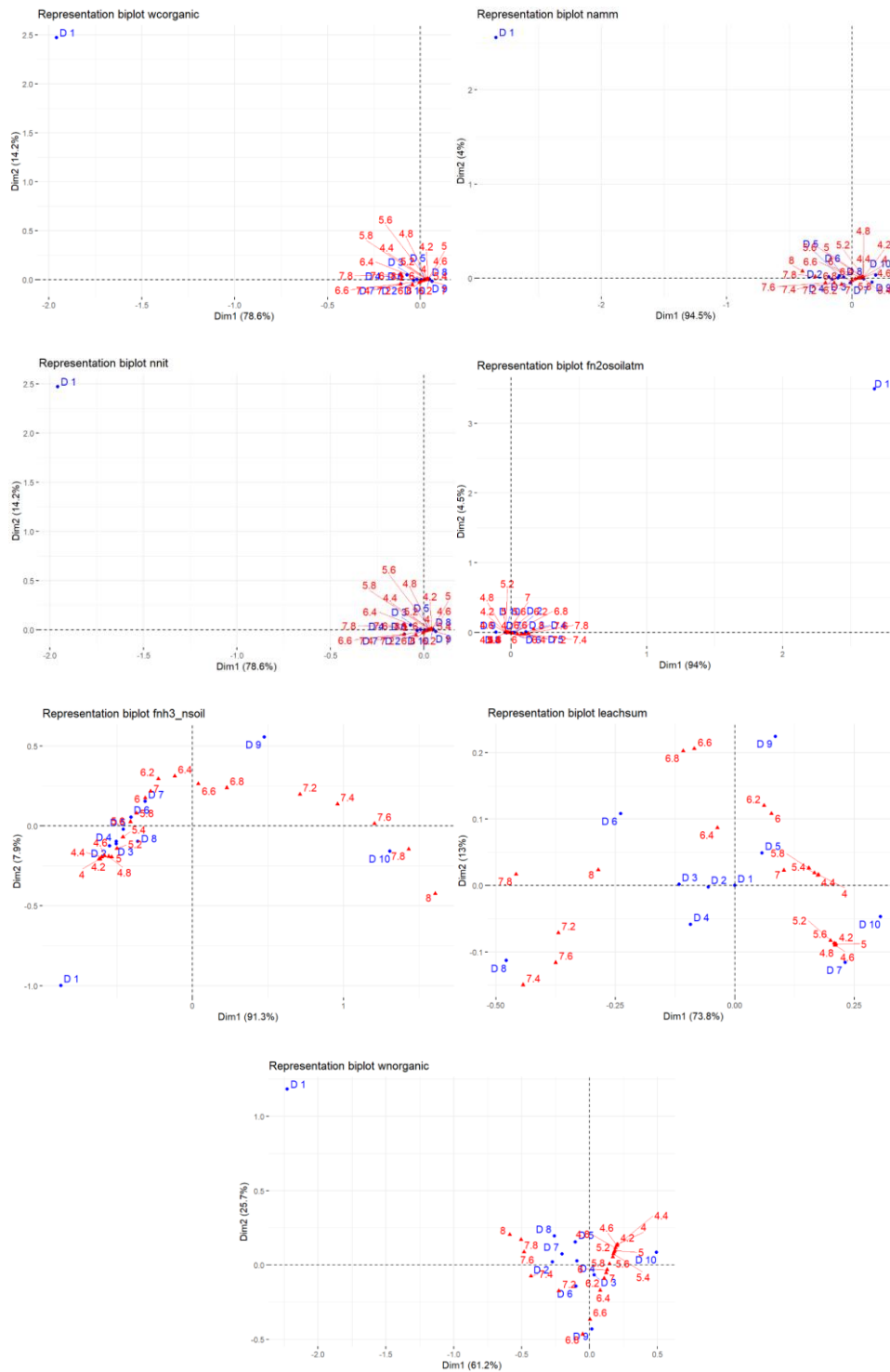
Quantité de nitrogène organique total dans le sol, en kgN m⁻²



Annexe n°8 : Biplot AFC - site de Laqueuille



Annexe n°9 : Biplot AFC - site de Theix fertilisation NPK



Annexe n°10 : Biplot AFC - site de Theix fertilisation NUL



Annexe n°11 : Présentation des méthodes d'analyse de sensibilité

Annexe n°11a : Méthodes de criblage

Dans les méthodes dites de criblage, l'objectif est de faire ressortir quelles entrées sont significatives sans chercher à être exhaustifs. Elles cherchent à isoler les paramètres les plus sensibles sur les sorties afin de réduire le nombre de paramètres à analyser. Ces méthodes exigent donc peu de simulations et sont très utiles pour les modèles de grande dimension car lors du traitement d'un problème impliquant un grand nombre d'entrées, il est raisonnable de penser que seuls quelques-unes d'entre elles sont réellement influentes.

Exemples de méthodes de criblage : plan factoriel complet, plan factoriel fractionné, plan de Morris et bifurcation séquentielle.

Méthodes OAT

Dans les méthodes dites OAT, un point de référence est considéré, parmi les valeurs pouvant être prises par les facteurs d'entrée. Puis une par une, la valeur de chaque paramètre est modifiée, afin d'analyser les effets que cela produit sur la sortie. Remarquer un changement à la sortie peut ainsi être facilement attribué au paramètre modifié.

Ces méthodes sont donc extrêmement répandues et très pratiques lorsque le modèle ne converge pas bien qu'elles ne permettent pas d'étudier l'effet de la variation de plusieurs variables d'entrée simultanément ni de détecter les possibles interactions entre les variables d'entrée.

Expérience factorielle à deux niveaux

Une application aux méthodes OAT est la théorie des expériences factorielles à deux niveaux (plans 2^k). Elle a été développée à la fin des années 1950 et au début des années 1960 et est encore en cours d'amélioration aujourd'hui. Cette théorie suppose que le modèle étudié a k paramètres qui sont définis comme ayant un comportement à deux niveaux différents. Ces niveaux sont appelés niveau inférieur et supérieur, ou respectivement niveaux -1 et $+1$ (pouvant être associé au maximum et au minimum pris par une variable). Les entrées du modèle ainsi définies, l'ensemble des différents croisements possibles entre ces facteurs sont simulés afin d'en analyser les sorties et les interactions qui en découlent.

Pour cela, il faudra pour chaque croisement calculer le produit scalaire $\hat{\theta}_i = \frac{1}{k/2} \mathbf{V}_i^T \mathbf{Y}$ avec \mathbf{V}_i le vecteur des niveaux choisis (Ex : $\mathbf{V} = (-1 \ 1 \ 1 \ -1)$) et k le nombre de paramètres étudiés, ou bien la différence entre la moyenne des sorties dont l'indice associé au facteur étudié est fixé à 1 et la moyenne des sorties dont l'indice associé au facteur étudié est fixé à -1 .

Pour déterminer si l'impact du facteur ou d'une interaction est pertinent ou non, il est possible d'utiliser l'écart moyen absolu (MAD) ou bien l'analyse de la variance (ANOVA). L'ANOVA est utilisée lorsque le nombre de facteurs et d'interactions dans le modèle statistique est supérieur au nombre d'observations. Dans le cas inverse, la MAD sera utilisée.

Néanmoins, ces tests sont basés sur le fait que les données proviennent d'expériences affectées par des perturbations aléatoires, ce qui n'est pas le cas des modèles informatiques. A cause de l'absence de variabilité expérimentale, les résultats des tests obtenus à partir de codes informatiques, doivent donc être considérés comme des indications, plutôt que comme de véritables tests statistiques.

Cette méthode peut être réalisée sur l'ensemble des entrées du modèle ou sur un plan factoriel fractionné. Dans ce cas, il faudra choisir quelles variables écarter et les fixées à -1 ou 1 , le nombre de cas considérés sera alors réduit de moitié. Cela est réalisable grâce au principe de la rareté des effets, c'est-à-dire que dans le cas où il y a beaucoup d'entrées, la sortie ne dépendra probablement pas de l'ensemble d'entre elles de manière significative. De plus ces expérimentations séquentielles peuvent être combinées afin d'obtenir des informations sur le système de plus en plus conséquentes et projetées dans des dessins plus grands dans le sous-ensemble des facteurs significatifs.

Méthode de Morris

La méthode de Morris (1991) est une méthode OAT et de criblage à coût modéré : bien que considérée comme globale, cette méthode est fondée sur le calcul des mesures locales. Cette méthode a pour but d'identifier les facteurs influents et en fonction de cela les classer. Les limites de cette méthode sont atteintes lors de la présence d'interactions d'effets non linéaires. Mais elle a l'avantage d'être économe grâce à son petit nombre d'appels au modèle.

La méthode repose sur le fait qu'en 1991, Morris a pensé que la dérivée partielle de Y (la sortie du modèle) par rapport à X_i (la $i^{\text{ème}}$ entrée) en un point x donné du plan de Y est une mesure de l'influence de X_i sur Y . Ainsi, Morris considère que si la dérivée partielle est :

- 0 sur toutes les valeurs de x . Alors l'effet de X_i sur Y est négligeable.
- une constante non nulle sur toutes les valeurs de x . Alors l'effet de X_i sur Y est linéaire et additif.
- une fonction non constante de X_i seulement. Alors l'effet de X_i sur Y est non linéaire.

- une fonction non constante d'un ou plusieurs X_j ($j \neq i$). L'effet de X_i sur Y implique alors des interactions avec d'autres facteurs d'entrée.

En application, la méthode de Morris nécessite un espace discret. C'est pourquoi, lors de sa réalisation, il faut dans un premier temps définir un plan d'expérience. Un élément de ce plan est alors choisi aléatoirement parmi l'ensemble des valeurs possibles afin d'être le point de départ de l'analyse. Ensuite, un saut, noté Δ_i , est ajouté au $i^{\text{ème}}$ facteur du point de départ, i étant un entier aléatoire entre 0 et n . La procédure est alors répétée pour tous les paramètres à r (nombre arbitraire) reprises et pour toutes répétitions, les effets élémentaires sont calculés pour à la fin de l'étude être analysés.

Afin d'estimer la sensibilité des n paramètres d'un modèle, la méthode de Morris consiste à effectuer $n+1$ répétitions à un nombre arbitraire de reprises. Pour chaque répétition, les effets élémentaires du paramètre i étudié noté EE_i sont calculées :

$$EE_i = \frac{dy}{dx_i} = \frac{y(x_1, x_2, x_3, x_i + \Delta, \dots, x_p) - y(x)}{\Delta}$$

avec y la sortie du modèle évalué à certaines valeurs nominales des paramètres, x le vecteur du point de référence, x_i le $i^{\text{ème}}$ paramètre du modèle et Δ le saut

Ainsi au cours des r répétitions, chaque paramètre sera associé à un effet élémentaire moyen

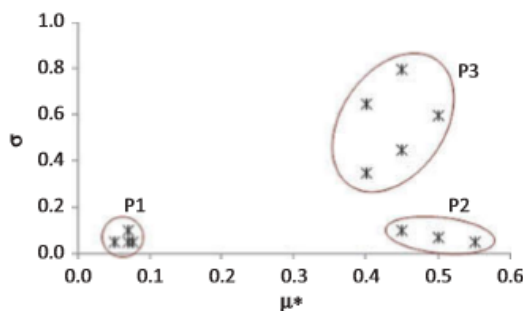
calculé ainsi : $\mu_i^* = \frac{1}{r} \sum_{j=1}^r |EE_i^j|$ et à un écart-type moyen étant égal à : $\sigma_i^2 = \frac{1}{r-1} \sum_{j=1}^r (EE_i^j - \mu_i)^2$

Un effet élémentaire moyen élevé révèle un paramètre influent : une variation de ce paramètre engendrera une variation importante sur le modèle. Un écart type faible permettra de considérer que l'effet de la variable sur le modèle est linéaire, tandis qu'à contrario, un écart type conséquent révélera un effet non-linéaire ou provoquant des interactions.

Ces résultats peuvent être représentés graphiquement avec μ_i en abscisse et σ_i en ordonnée pour chaque paramètre x_i . Ainsi, des conclusions pourront être tirées :

- points proches de l'origine (0,0) : facteurs dont les effets sont négligeables
- points situés à droite sur l'axe des abscisses : facteurs dont l'effet linéaire est important
- points situés en haut sur l'axe des ordonnées : facteurs dont les effets sont non-linéaires ou incluent de l'interaction avec d'autres facteurs

C'est pourquoi, dans l'exemple de la figure suivante les paramètres P1 ont des effets négligeables, les effets des paramètres P2 sont linéaires et les paramètres P3 ont des effets non linéaires ou avec des interactions



Exemple du graphique d'interprétation des indices de Morris (adapté de Saltelli et al. 2004)

Bifurcation séquentielle

La méthode de la bifurcation séquentielle a été développée par Bettonvil et Kleijnen (1996) et est une technique de criblage de groupe. Elle s'effectue dans un plan factoriel à 2 niveaux et permet de sélectionner un grand nombre de facteurs. L'un des principaux avantages est que le nombre de simulations à effectuer est inférieur au nombre de facteurs à examiner. L'inconvénient est que la relation entre chaque facteur d'entrée et la sortie doit être monotone et que le type de monotonie doit être connu a priori. Lorsque cette condition est remplie, la méthode est très efficace.

La méthode consiste à diviser en deux sous-groupes les variables et de sélectionner à chaque itération le groupe le plus influent. Ainsi, les facteurs les plus influents pourront être identifiés. Cette méthode peut s'effectuer dans un modèle linéaire avec ou sans interactions.

Ecriture d'un modèle linéaire sans interaction : $Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$

Ecriture d'un modèle linéaire avec interaction : $Y = \beta_0 + \sum_{i=1}^k \beta_i X_i + \sum_{\substack{i,j=1 \\ i < j}}^k \gamma_{ij} X_i X_j$

avec les facteurs β_i et γ_{ij} connu

L'effet du groupe des facteurs inclus entre le $i^{\text{ème}}$ et le $j^{\text{ème}}$ facteur se note e_{ij} et est défini comme la différence entre la moyenne des sorties Y avec les facteurs $X_i \dots X_j$ fixé à 1 et la moyenne des sorties Y avec les facteurs $X_i \dots X_j$ fixé à -1.

Il faudra alors 2 lancers pour estimer les effets d'un groupe dans un modèle sans interaction et 4 seront nécessaires pour un modèle avec interaction. La procédure étant itérative, le nombre de simulation diminue. A chaque itération, un groupe dont les effets sont déjà estimés est scindé en 2. C'est pourquoi, le nombre maximal de procédures est de $k+1$ pour le modèle linéaire et de $2k$ pour le modèle linéaire avec interactions. En pratique, ce coût de calcul est encore plus faible, car les itérations doivent s'arrêter lorsque tous les paramètres influents ont été trouvés.

Annexe n°11b : Méthodes graphiques

Les méthodes graphiques permettent de soutenir, guider et interpréter les résultats fournis par les techniques d'analyse de sensibilité numérique mais ce sont également des techniques autonomes permettant d'obtenir des informations supplémentaires sur le modèle étudié.

Diagrammes de dispersion

Tracer les diagrammes de dispersion de la variable de sortie en fonction de variables d'entrée suite à l'évaluation du modèle sur un échantillon aléatoire de données permet de donner une indication visuelle de la sensibilité.

Pour une entrée donnée X_i , le diagramme de dispersion est la projection des points d'échantillonnage (X, Y) sur le plan (X_i, Y) . Cette représentation permet d'identifier les relations linéaires, les relations monotones et l'existence de seuils. Pour aider à tirer des conclusions de ces diagrammes, des transformations telles que les logarithmes ou les rangs peuvent être utilisées. L'utilisation des rangs est applicable pour n'importe quelles valeurs alors qu'il faut veiller à n'avoir aucune valeur qui s'annule pour utiliser un logarithme. Cependant le logarithme semble être une alternative plus intuitive que les rangs. En effet, il permet de regrouper les valeurs afin que celles-ci soient plus facilement représentables tout en conservant la notion de dispersion (si une valeur était très

éloignée des autres, elle le restera) alors que cela n'est pas le cas pour les rangs qui considèrent tous les points à équidistance les uns des autres (deux valeurs se succédant mais étant séparées par plusieurs ordres de grandeur auront un rang associé de n et $n+1$ comme ce serait le cas pour deux valeurs très proches).

Ainsi pour analyser la sensibilité d'une sortie à une variable il est envisageable de créer les diagrammes de dispersion avec les valeurs brutes, leur logarithme et leur rang afin d'étudier la dépendance entrée-sortie, la répartition de la partie non nulle de la sortie sur 10 ordres de grandeur, le rang de la sortie par rapport au rang de l'entrée ou la présence de seuil.

Les diagrammes de dispersion peuvent également être utilisés pour étudier la dépendance entre les variables de sortie. Bien que cela ne puisse pas être considéré comme une véritable analyse de sensibilité, cela peut fournir des informations très utiles sur le modèle du système. Cela se remarque lorsque certains points sont proches de la droite $y=x$.

Pour que ces diagrammes soient lisibles sur un écran cela peut se faire en 2 voir 3 dimensions c'est-à-dire en ne s'intéressant qu'à 1 ou 2 variables d'entrées (la dernière dimension correspondant à celle de sortie).

Les extensions des diagrammes de dispersion sont des matrices de diagrammes de dispersion ou des superpositions de diagrammes de dispersion afin de simuler simultanément les diagrammes de dispersion de différentes paires de paramètres d'entrée/variables de sortie.

Diagrammes en toile d'araignée

Les diagrammes en toile d'araignée ont été conçus pour montrer des échantillons multidimensionnels dans un graphique en 2 dimensions, cette méthode a en particulier été mise en évidence par Cooke et Van Noortwijk en 1999.

Cette représentation est utilisée lors de l'étude d'une dizaine de variables maximum par soucis de lisibilité. Cela consiste à associer chaque variable d'entrée étudiée ainsi que la sortie du modèle à une ligne verticale. L'axe des abscisses du graphique est ainsi divisé en plusieurs parties correspondants aux nombres de variables observées. A chaque exécution, les valeurs de chaque variable sont alors fixées sur la ligne qui lui est associée puis les valeurs d'un même échantillon sont reliées par (en règle générale) une ligne de couleur. Il est nécessaire de faire attention aux valeurs considérées, il est possible d'utiliser les rangs ou bien les valeurs brutes mais il n'est pas judicieux de mélanger les deux. Alors, suite à l'étude de ces diagrammes, des conclusions sur la dépendance entre certaines variables peuvent alors être déterminées ainsi que sur la complexité de la relation de la sortie avec les entrées.

D'autres conclusions peuvent aussi être mises en évidence lors de l'utilisation de parcelles de toile d'araignée conditionnelle c'est-à-dire en ne représentant qu'une partie des exécutions comme les 10% renvoyant les sorties les plus hautes par exemple.

Diagrammes de moyenne d'échantillon (Anca Badea and Ricardo Bolado)

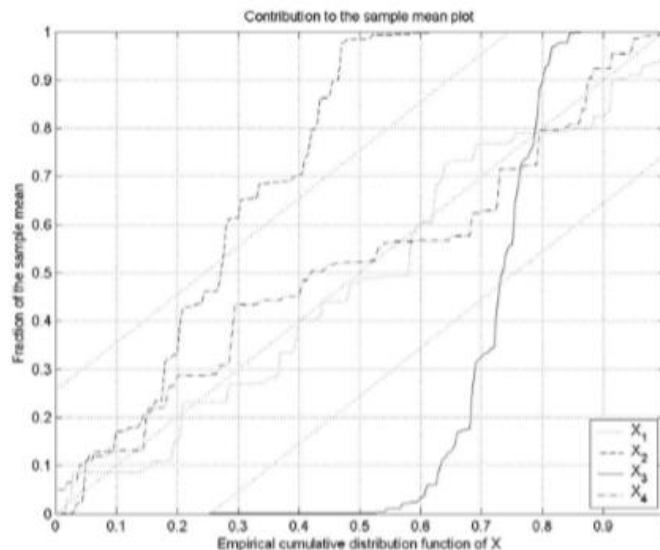
Les diagrammes de moyenne d'échantillon ont été développés en 1993 par Sinclair qui a étudié la façon dont de minuscules modifications de la fonction de densité de probabilité d'une variable d'entrée X_i peuvent modifier la sortie Y . Bien que la dépendance marginale de $E(Y)$ par rapport aux différents facteurs d'entrée puisse être utilisée et représentée graphiquement.

Cette méthode n'est applicable que lorsque les valeurs de la sortie Y sont positives ou nulle et en supposant qu'un échantillon aléatoire S de taille n des facteurs d'entrée et sa sortie associée sont disponibles. Cet échantillon aléatoire doit avoir été obtenu par une technique d'échantillonnage qui n'introduit aucun biais (échantillonnage aléatoire simple, échantillonnage stratifié proportionnel, ...), sans quoi, la diagonale qui sera utilisé pour l'interprétation du graphique ne sera plus une référence pour mesurer le manque d'importance d'une variable d'entrée et l'interprétation de la représentation graphique deviendra difficile.

Pour réaliser le tracé du diagramme de moyenne d'échantillon pour une variable d'entrée X_i et la réponse Y il faut donc :

- Trier les réalisations de X_i en générant la série de valeurs $\{x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(n)}\}$
- Créer la série de sorties correspondantes $\{y^{[1]}, y^{[2]}, \dots, y^{[n]}\}$:
- Calculer la variable auxiliaire associée à chaque série $m_i^q = \frac{1}{n} \sum_{j=1}^q y^{[j]}$ avec q un entier variant de 1 à n
- Normaliser les variables auxiliaires en divisant la valeur m_i^q par la moyenne de l'échantillon Y
- Tracer les valeurs échantillonnées de M_i en fonction de la distribution cumulative de X_i .

Chaque point d'un tracé de contribution à la moyenne représente la fraction de contribution de la moyenne sur la sortie en fonction de la distribution cumulative empirique de x. C'est indirectement une contribution à la variance. Lors de cette représentation, la diagonale et l'intervalle de confiance de 95% autour de celle-ci sont fréquemment inclus afin de faciliter la lecture du graphique.



Graphique de la contribution à la moyenne pour quatre paramètres d'entrée et une variable de sortie

Dans la figure n°6, il est possible de mettre en évidence que :

- Lorsque le tracé de la contribution à la moyenne de l'échantillon d'une variable est proche de la diagonale, la variable a toujours le même impact sur la moyenne de l'échantillon de sortie. (X_1, X_4)

- Un tracé ressemblant à une droite avec un coefficient de corrélation élevée pour une région de valeurs prises par la fonction de distribution cumulative est témoin d'une contribution importante de ces valeurs sur la moyenne de la sortie (pour X_2 60% des plus petites valeurs sont responsables de plus de 99% de la moyenne de l'échantillon de sortie, pour X_3 seules les valeurs entre les 50% et les 85% sont responsables de la moyenne de l'échantillon de sortie).

Ainsi, l'importance d'une variable peut être déterminée. Une variable est considérée comme importante si connaître la valeur prise par cette variable permet de fournir des informations pertinentes sur la sortie. C'est le cas lorsque graphiquement il est possible d'observer des écarts nets par rapport à la diagonale.

Ces graphiques permettent donc de représenter dans une même figure la relation de nombreux paramètres d'entrée avec une variable de sortie et ils peuvent aussi être utilisés pour étudier la relation entre un paramètre d'entrée et de nombreuses variables de sortie, voire pour représenter la relation entre différentes entrées et sorties.

Annexe n°11c : Méthodes globales – A Badea et Ricardo Bolado

Les méthodes globales se concentrent sur la façon dont l'espace d'entrée s'intègre dans l'espace de sortie.

Exemples de méthodes globales : les méthodes de Monte Carlo, les méthodes basées sur la décomposition de la variance et les méthodes de sensibilité de la distribution. Dans les pages suivantes, les techniques les plus intéressantes sont décrites.

Méthode de Monte Carlo

La méthode de Monte Carlo est la plus simple et la plus puissante de la littérature scientifique. Elle permet de traiter la propagation de l'incertitude dans des modèles complexes. Pour cela, la méthode consiste à échantillonner aléatoirement le vecteur des paramètres d'entrée, puis exécuter le modèle pour chaque échantillon afin d'obtenir un échantillon de sorties qu'il sera alors possible d'analyser afin d'en faire ressortir certaines caractéristiques.

Cette méthode est applicable avec des entrées et des sorties discrètes et continues ainsi que pour les modèles avec des sorties statiques et dynamiques. De plus, cette méthode est souvent utilisée car les algorithmes de calcul nécessaires sont peu complexes.

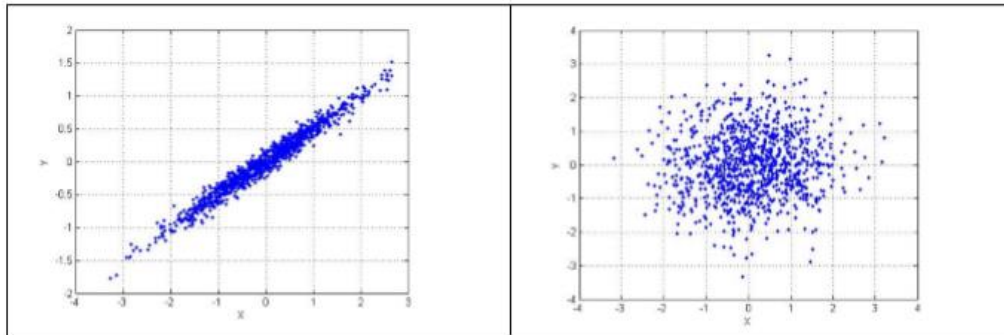
Avec cette méthode, l'analyse la plus couramment utilisée est celle basée sur le calcul du coefficient de corrélation. Le coefficient de corrélation fournit une mesure de la force de la relation linéaire entre tout facteur d'entrée X_j et la sortie Y , il se calcule grâce au coefficient de Pearson (le coefficient de régression standardisé d'une régression), le coefficient de Spearman ou par la formule :

$$CC(x_j, y) = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad \text{avec} \quad \bar{x}_j = \sum_{i=1}^n x_{ij} / n \quad \text{et} \quad \bar{y} = \sum_{i=1}^n y_i / n \quad \text{les moyennes des échantillons}$$

Un coefficient de corrélation positif entre X_i et Y , implique que lorsque X_i augmente Y augmente également (et si X_i diminue Y diminue). A l'inverse, un coefficient négatif implique qu'une augmentation de X_i engendre une diminution de Y (et si X_i diminue Y augmente). Ces coefficients

oscillent entre -1 et 1 plus le coefficient sera proche de la valeur 1 (ou -1) plus la corrélation entre X_i et Y sera forte et plus le coefficient sera proche de 0 plus la corrélation entre la variable d'entrée et celle de sortie sera faible.

Graphiquement, une corrélation positive forte se traduit par des valeurs semblant former une droite (figure de gauche, $CC = 0,9825$) tandis qu'une corrélation faible se traduit par un nuage de point qui semble être aléatoire (figure de droite, $CC = 0,0699$).



Echantillon avec une forte relation linéaire (gauche) et aucune relation linéaire (droite) entre X_i et Y

Dans le cas d'une régression linéaire, l'utilisation du coefficient de corrélation de Pearson est préférable. Lorsqu'il existe une relation monotone entre X_j et Y , il faut effectuer une transformation de rang afin d'obtenir une relation linéaire et ainsi pouvoir utiliser le coefficient de corrélation de Spearman. Il est calculé en utilisant la même expression que le coefficient de corrélation de Pearson, le seul changement est l'utilisation de rang au dépend de valeurs brutes.

En programmation, le coefficient de détermination multiple est renvoyé automatiquement suite à une régression. Il correspond au carré du coefficient de corrélation et permet de déterminer à quel point la régression faite sur le modèle correspond ou non aux données et explique ses incertitudes. (la majeure partie de l'incertitude en y est prise en compte lorsque R^2 est proche de 1).

Les coefficients de régression standardisés (SRC) peuvent également être utilisés. Ils fournissent une mesure de sensibilité, ou plus précisément une mesure de l'importance des variables : une variable X_j est plus importante qu'une variable X_i si la valeur absolue du coefficient de régression normalisé de X_j est plus grande que la valeur absolue du coefficient de régression normalisé de X_i .

Lorsque la régression n'est ni linéaire ni monotone, il est possible qu'une entrée et une sortie aient une corrélation forte s'il existe un lien clair entre des régions spécifiques de chacune des variables. Dans ces cas-ci, le filtrage de Monte Carlo peut être mis en place : le vecteur Y est trié dans l'ordre croissant, le vecteur X est ré-ordonné en fonction de cet arrangement (ainsi les vecteurs $(y^{(1)}, y^{(2)}, \dots, y^{(n)})$ et $(x^{[1]}, x^{[2]}, \dots, x^{[n]})$ sont obtenus). L'échantillon de sortie est alors divisé en plusieurs sous-ensembles en fonction d'un critère particulier (une condition, un seuil, ...) ainsi des tests sont réalisées sur les différences existantes ou non entre les entrées associées à ces sous-ensembles tels que :

- Le test de Smirnov à deux ou k échantillons : basé sur la comparaison de fonctions de distribution cumulées empiriques

- Le test de Mann-Whitney (ou Wilcoxon) à deux échantillons et le test de Kruskal-Wallis (par extension) : basé sur les rangs

- Le t-test à deux échantillons : basé sur la distribution d'échantillonnage de la moyenne des variables normales

Méthodes sur la décomposition de la variance

Ces méthodes permettent d'explorer l'ensemble de l'espace des entrées en prenant en compte les interactions ainsi que les réponses non-linéaires. Contrairement aux méthodes OAT, elles calculent les effets de chaque facteur d'entrées en les faisant varier simultanément. Pour cela, elles considèrent les entrées comme des variables aléatoires et s'intéressent seulement à la variance de la sortie, puis les indices de Sobol sont calculés. Ce sont des indices de sensibilité qui représentent la proportion de variance expliquée par une entrée ou un groupe d'entrées. Si est l'indice de sensibilité de premier ordre du $i^{\text{ème}}$ paramètre. Il permet de quantifier la part de la variance de Y causé par le paramètre X_i et est

défini par : $S_i = \frac{\text{Var}[\mathbf{E}[Y|X_i]]}{\text{Var}[Y]}$. Tandis que ST_i est le $i^{\text{ème}}$ indice de sensibilité totale défini comme la somme de tous les indices de sensibilité impliquant le paramètre X_i , il prend également en compte les interactions du paramètre X_i avec les autres paramètres contrairement à l'indice précédent et se définit

comme étant égal à : $ST_i = 1 - \frac{\text{Var}[\mathbf{E}[Y|X_{\sim i}]]}{\text{Var}[Y]}$

Pour analyser ces indices, il faut retenir que pour un indice i donné :

- S_i grand : forte influence de X_i (à lui seul) sur Y
- ST_i petit : peu d'influence de X_i (et de ses interactions avec les autres paramètres) sur Y
- $S_i = ST_i$: pas d'influences des interactions de X_i avec les autres entrées sur Y
- S_i faible mais ST_i élevé : impact sur la sortie dû à ses interactions.
- Pas d'effet individuel ni d'effet d'interactions avec les autres paramètres : paramètre non influent et pouvant être fixé à n'importe quelle valeur dans sa gamme d'incertitude.
- De plus, dire que deux paramètres interagissent entre eux revient à dire que leur effet total sur la sortie n'est pas égal à la somme de leurs effets du premier ordre. Un modèle sans interaction est donc dit additif (exemple : modèle linéaire).

Cependant, le calcul des variances des attentes conditionnelles nécessaires au calcul des indices de Sobol prend beaucoup de temps. C'est pourquoi il est nécessaire d'utiliser une méthode telle que :

- La méthode de Sobol/Saltelli,
- La méthode FAST (Fourier Amplitude Sensitivity Test), la méthode de FAST étendu ou la méthode de conception aléatoire et équilibrée
- L'utilisation des ratios de corrélation
- La méthode de McKay

Une autre possibilité moins répandue consiste à construire une surface de réponse pour le modèle initial et à calculer ou estimer les indices de Sobol pour cette surface de réponse.

Enfin la méthode Pick and Freeze est une alternative à aux méthodes de décomposition de la variance lorsque les entrées ne sont pas indépendantes.

Méthode d'estimation de Sobol et Saltelli – Pick and Freeze

Sobol a créé une méthode d'estimation des indices en 1993 puis elle a été développée par Saltelli et al. en 2000. Grâce à Saltelli, le nombre d'évaluations du modèle nécessaire pour estimer tous les indices de premier ordre et totaux a pu être réduit à $n(d+2)$ (avec d le nombre d'indices étudiés).

Cette approche se nomme la méthode Pick and Freeze. Elle est très adaptée aux modèles fonctionnels multidimensionnels. Elle ne nécessite aucune condition sur le nombre de variables par lequel la variance est conditionnée. La méthode Pick and Freeze est l'approche initialement proposée par Sobol pour estimer les indices de sensibilité qui reposent sur les moments d'ordre 1 et 2 et peuvent être calculés à l'aide de simulations de type Monte Carlo de $X(i)$ sous la forme $(X_1(i), \dots, X_p(i))_{i=1, \dots, N}$ et la mesure des sorties correspondantes $(Y(i))_{i=1, \dots, N}$. Ainsi en posant $X = (U, V)$ et $Y = \eta(U, V)$ avec $U = \{X^{i_1}, \dots, X^{i_k}\}$ le vecteur des k facteurs d'entrée, V le vecteur des $d-k$ facteurs complémentaires et η la fonction simulant la relation entrée/sortie, l'indice de Sobol sera défini comme étant égal à :

$$S^U = \frac{\text{Cov}(Y^U, Y)}{\text{Var}(Y)} .$$

En considérant un échantillon de taille N , un estimateur Pick and Freeze sera le coefficient de corrélation entre la sortie du modèle et sa réplique "pick-frozen". Cette réplique est obtenue en retenant la valeur de la variable d'intérêt (variable gelée) et en échantillonnant les autres variables (variables choisies). Les répliques échantillonnées sont ensuite combinées pour produire un estimateur de l'indice de Sobol. L'estimateur vaudra alors :

$$\widehat{S}_N^U = \frac{\frac{1}{N} \sum_{i=1}^N Y^{(i)} Y^{U,(i)} - \left(\frac{1}{N} \sum_{i=1}^N Y^{(i)}\right) \left(\frac{1}{N} \sum_{i=1}^N Y^{U,(i)}\right)}{\frac{1}{N} \sum_{i=1}^N (Y^{(i)})^2 - \left(\frac{1}{N} \sum_{i=1}^N Y^{(i)}\right)^2}$$

Les indices de sensibilité de premier et de second ordre ainsi que leur intervalle de confiance à 95% sont ensuite représentés graphiquement afin d'être analysés.

Cette méthode est robuste et sa vitesse de convergence ne dépend pas du nombre de variables d'entrée.

Méthode FAST - Fourier

La méthode FAST a été initialement créée par Cukier et al. en 1973 puis elle a été améliorée à de nombreuses reprises. La méthode FAST est très semblable à la méthode de Sobol mais elles se distinguent par leur approche du calcul des intégrales multidimensionnelles. Le calcul des intégrales de la méthode FAST repose sur la transformation de Fourier multidimensionnelle d'une fonction pour obtenir la décomposition de la variance, tandis que celui de la méthode Sobol utilise la méthode de Monte Carlo.

Cette méthode est basée sur le fait qu'il est possible de convertir les intégrales k -dimensionnelles définissant la décomposition de la variance en intégrales unidimensionnelles, en utilisant les fonctions de transformation G_i avec i compris entre 1 et k . Ainsi G_i est définie comme :

$x_i = G_i(\sin(\omega_i s))$ où s est une variable scalaire comprise entre $-\pi$ et π et $\{\omega_i\}$ est un ensemble de fréquences angulaires entières.

Ainsi en utilisant les propriétés des séries de Fourier, l'espérance et la variance de Y peuvent

$$\begin{aligned} \text{Var}(Y) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f^2(s) ds - f_0^2 \approx \sum_{j=-\infty}^{\infty} (A_j^2 + B_j^2) - (A_0^2 + B_0^2) \\ \text{être approximées par : } f_0 &= E[Y] = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) ds \quad \text{et} \quad = 2 \sum_{j=1}^{\infty} (A_j^2 + B_j^2) \quad \text{avec} \\ A_j \text{ et } B_j, \text{ les coefficients de Fourier} \quad A_j &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) \cos(js) ds, \quad B_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) \sin(js) ds. \end{aligned}$$

Cukier et al. ont alors montré que la part de variance d'une entrée X_i est la somme des carrés des coefficients de Fourier due à la fréquence angulaire ω_j et ses harmoniques vaut :

$$V_i = 2 \sum_{k=1}^{\infty} (A_{k\omega_i}^2 + B_{k\omega_i}^2).$$

L'indice de sensibilité S_i peut alors se définir ainsi

$$S_i = \frac{\sum_{k=1}^{\infty} (A_{k\omega_i}^2 + B_{k\omega_i}^2)}{\sum_{j=1}^{\infty} (A_j^2 + B_j^2)}.$$

Ainsi calculés, les indices de sensibilité de premier et de second ordre et leur intervalle de confiance à 95% pourront comme précédemment être représentés graphiquement afin d'être analysés.

Méthode d'estimation EFAST - Anca Badea et Ricardo Bolado

Cette méthode est une extension de la méthode précédente, développée par Saltelli et al. en 1999 où l'indice considéré est la partie de la variance due à toutes les variables autre que X_i , soit :

$$S_{T_i} = 1 - \frac{\sum_{k=1}^{\infty} (A_{k\omega_i}^2 + B_{k\omega_i}^2)}{\sum_{j=1}^{\infty} (A_j^2 + B_j^2)}$$

De plus, dans cette méthode, les fonctions G_i sont calculées d'une autre façon afin de mieux s'intégrer dans l'espace d'entrée, elles sont alors définies ainsi :

$$\begin{aligned} x_i &= G_i(\sin(\omega_i s)) = \frac{1}{2} + \frac{1}{\pi} \arcsin(\sin(\omega_i s)) \\ x_i &= G_i(\sin(\omega_i s)) = \frac{1}{2} + \frac{1}{\pi} \arcsin(\sin(\omega_i s + \varphi_i)) \end{aligned}$$

avec φ_i une phase aléatoire choisie uniformément dans $[0, 2\pi[$ et ω_i le choix des fréquences angulaires qui n'est plus contraint à l'absence d'interférences jusqu'à un certain ordre mais qui doit tout de même être grande par rapport aux fréquences angulaires correspondant au reste des variables. Suite à ces calculs, comme pour la méthode FAST classique, ces échantillons sont représentés graphiquement puis analysés.

La méthode FAST (ou EFAST) est dite déterministe, il n'est donc pas possible de calculer les intervalles de tolérance (contrairement à la méthode Sobol). Pour estimer $2*d$ indices de sensibilité, il est nécessaire d'effectuer $n*d$ exécutions.

Estimation par plan équilibré aléatoire (RBD) – Tarantola et al. (2006), Saltelli et al. (2008)

La méthode RBD est similaire aux méthodes FAST ou EFAST sauf pour l'exploration de l'espace d'entrée où la fréquence ω utilisée reste toujours la même, pour limiter l'utilisation d'algorithmes complexes lors du choix des différentes fréquences. Ainsi, l'ensemble des paramètres d'entrée ne remplit plus l'espace, les permutations aléatoires des coordonnées de ces points sont alors prises en compte afin de générer un ensemble de points brouillés qui couvrent l'espace d'entrée. Le modèle est ensuite évalué à chaque point de conception. Pour calculer l'indice de sensibilité du premier ordre S_i , la sortie y est réordonnée dans l'ordre croissant des valeurs prises par le facteur X_i . Le spectre de Fourier de la sortie est calculé pour les fréquences $\{\omega, 2\omega, \dots, M\omega\}$ afin d'estimer l'indice de sensibilité S_i . La procédure est répétée pour tous les autres facteurs afin d'obtenir tous les indices de sensibilité de premier ordre.

Contrairement à FAST et EFAST, il n'y a pas de taille minimale d'échantillon pour éviter les problèmes d'interférence et aucun algorithme spécial n'est nécessaire pour fournir l'ensemble de fréquences à interférence libre. De plus, les intervalles de confiance de ces indices peuvent être calculés par cette méthode. Cependant elle ne permet de calculer uniquement les indices de sensibilité de premier ordre.

Pour estimer d indices de sensibilité avec cette méthode, il est nécessaire d'effectuer n exécutions.

Méthodes de sensibilité de la distribution

De nombreuses incertitudes apparaissent lors de l'étude de certains modèles, les distributions fournies lors de cette étude sont peu précises il est donc nécessaire d'en mesurer l'impact sur la distribution des résultats. Cela peut se faire grâce aux méthodes de sensibilité de la distribution qui permettent de mesurer l'influence des changements de distribution des entrées dans les moyennes et les fonctions de distribution des variables de sortie.

Pour mettre en place ces méthodes, il est nécessaire d'exécuter le code n fois sous la distribution originale, puis n' fois sous la distribution alternative, afin de comparer les résultats (moyennes, fonctions de distribution, probabilités de succès ou d'échec). Lorsque l'étude est basée sur k paramètres d'entrée, le code devra être exécuté $n+kn'$ fois pour mesurer l'influence de la modification de la distribution d'entrée de chaque paramètre d'entrée. Ces méthodes peuvent permettre de mesurer l'effet de la modification simultanée de la distribution de deux ou plusieurs paramètres d'entrée mais elles ne sont pas efficaces pour mesurer l'effet des changements de distribution des entrées sur la distribution des sorties.

Exemples de méthodes de sensibilité de la distribution : méthode de pondération, méthode de rejet, analyse de régression et ANOVA.

Méthode de pondération - Beckman et McKay 1987

La méthode de pondération convient pour mesurer l'effet sur la moyenne de la sortie et ressemble à la technique de l'échantillonnage par importance. Dans le cas de l'échantillonnage par importance, le problème consiste à estimer la moyenne d'une variable aléatoire de référence aussi précisément que possible, et la solution proposée consiste à trouver une variable aléatoire auxiliaire ayant la même moyenne et une variance plus faible que la variable de référence. Cette méthode n'est adéquate que pour l'estimation des variations de la moyenne et s'effectue ainsi :

1. Obtenir un échantillon de taille n du vecteur aléatoire d'entrée : X_1, X_2, \dots, X_n . Leurs valeurs réelles seront x_1, x_2, \dots, x_n .
2. Exécuter le code pour ces n ensembles d'entrées afin d'obtenir l'échantillon correspondant de la variable de sortie : Y_1, Y_2, \dots, Y_n . Leurs valeurs réelles seront y_1, y_2, \dots, y_n .
3. Pondérer l'échantillon avec les poids correspondants, en obtenant pour chaque valeur effectivement échantillonnée $y_i = y(x_i)$ la valeur correspondante

$$y'(x_i) = y(x_i) \omega(x_i) = y(x_i) f_2(x_i) / f_1(x_i),$$

avec f_1 la distribution de référence et f_2 la distribution de sensibilité

4. Estimation de $\hat{\mu}_2 = E_2(Y(X))$ grâce à :
$$\hat{\mu}_2 = \frac{1}{n} \sum_{i=1}^n \frac{f_2(\mathbf{x}_i)}{f_1(\mathbf{x}_i)} y(\mathbf{x}_i) .$$

Puis il suffit de comparer la précision de l'estimateur $\hat{\mu}_1$ de la moyenne de Y sous la distribution de référence avec la précision de l'estimateur $\hat{\mu}_2$ de la moyenne de Y sous la distribution de sensibilité. L'efficacité du second estimateur par rapport au premier sera défini comme le quotient de la variance du premier par la variance du second et sera donc égale au quotient des variances de Y et Y' .

Méthode de rejet - Beckman et McKay 1987

La méthode de rejet convient pour mesurer l'effet sur la fonction de distribution et est basée sur la technique d'échantillonnage par acceptation/rejet pour échantillonner à partir de distributions. Pour estimer la variation de la probabilité de succès d'un système passif, cette méthode est à privilégier.

L'application de cette méthode est soumise à deux conditions : Premièrement, comme dans la méthode précédente, le support de $f_2(x)$, R_2 , doit être contenu dans le support de $f_1(x)$, R_1 ; et deuxièmement, le quotient $f_2(x)/f_1(x)$ doit être limité. Supposons que la limite est M ; en d'autres termes, $f_2(x)/f_1(x) \leq M$. Alors la procédure à suivre est la suivante :

1. Obtenir un échantillon de taille n du vecteur aléatoire d'entrée sous la distribution de référence : X_1, X_2, \dots, X_n . Leurs valeurs réelles seront x_1, x_2, \dots, x_n .
2. Exécuter le code pour ces n ensembles d'entrées afin d'obtenir l'échantillon correspondant de la variable de sortie : Y_1, Y_2, \dots, Y_n . Leurs valeurs réelles seront y_1, y_2, \dots, y_n .
3. Pour chaque échantillon x_i , prélever un échantillon de la distribution uniforme V_i entre 0 et $M_{f1}(x_i)$.

4. Conserver dans l'échantillon la valeur de sortie correspondante $Y_i = y_i$ si la réalisation v_i de V_i est inférieure ou égale à $f_2(x_i)$, sinon rejeter cette valeur de l'échantillon.
5. Considérer les valeurs de Y restantes dans l'échantillon comme un échantillon aléatoire de Y sous $f_2(x)$, et utiliser cet échantillon de taille $k \leq n$ pour construire une fonction de distribution empirique qui estime la fonction de distribution réelle de Y sous $f_2(x)$. Chaque étape de cette fonction de distribution empirique aura une hauteur de $1/k$.

L'objectif est d'obtenir un estimateur de la probabilité d'échec du modèle sous $f_1(x)$ et $f_2(x)$. En supposant que le système est en situation d'échec lorsque la variable Y est au-dessus d'une valeur y_0 , l'estimateur sera le nombre d'observations de Y qui dépassent ce seuil divisé par n pour le cas de référence ($\hat{\mu}_1$) et par k pour le cas de sensibilité ($\hat{\mu}_2$). Cette méthode permet de produire un estimateur non biaisé.

Analyse de régression

Les coefficients de régression standardisés peuvent être utilisés comme des mesures directes de la sensibilité. Cette méthode est applicable lorsque la régression est linéaire par rapport aux données sans quoi il est difficile d'interpréter les coefficients standardisés. Cette méthode est simple et avec un faible coût de calcul.

ANOVA

La méthode de l'analyse de la variance (ANOVA) a pour objectif d'étudier la variation de la moyenne d'un phénomène étudié sous l'influence d'un ou plusieurs facteurs qualitatifs. Elle ne s'applique uniquement lorsque certaines conditions sont présentes :

- observations obtenues de manière indépendante et aléatoire
- données de chaque niveau de facteur normalement distribuées
- variance commune aux populations (vérification par le test de Levene)

L'ANOVA à un facteur est une extension du t-test à deux échantillons indépendants pour comparer les moyennes dans une situation où il y a plus de deux groupes. Dans l'ANOVA à un facteur, les données sont organisées en plusieurs groupes sur la base d'une seule variable de regroupement. Les hypothèses de ce test sont définies ainsi :

- H_0 (hypothèse nulle) : les moyennes des différents groupes sont les mêmes
- H_1 (hypothèse alternative) : au moins une moyenne de l'échantillon n'est pas égale aux autres.

Dans le cas d'une ANOVA à deux facteurs, il faut supposer qu'il n'y a pas de différence dans les moyennes du premier facteur, ni dans celles du second facteur, et qu'il n'y a pas d'interactions entre ces deux facteurs.

L'interprétation des résultats de l'ANOVA se fait grâce aux p-value avec un niveau de signification de 0,05, c'est-à-dire que lorsque la p-value d'une variable est inférieure à 0.05, les niveaux de cette variable sont associés à des sorties significativement différentes. Autrement dit, si la p-value est inférieure à 0.05, l'hypothèse nulle peut être rejetée.

Annexe n°11d : Méthodes locales - Dan G. Cacuci

Les méthodes locales se basent sur l'étude de la dérivée partielle de la sortie Y par rapport à une entrée X_i en un point fixe. Bien que proches des méthodes OAT, les méthodes locales examinent les petites perturbations (souvent une variable à la fois) et n'explorent pas nécessairement l'ensemble des entrées ni des sorties.

Exemples de méthodes locales : la procédure d'analyse de sensibilité prospective et la procédure d'analyse de sensibilité adjointe.

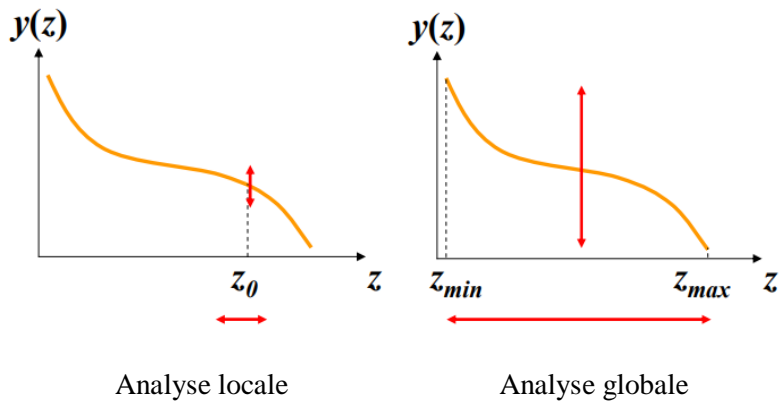
Le but de l'analyse de sensibilité est de calculer avec précision et efficacité les sensibilités de la réponse du système aux variations des paramètres autour de leur valeurs nominales. Dans le cadre d'une analyse de sensibilité locale, cela consiste à calculer la première différentielle de Gâteaux de la réponse du système. Cette différentielle au point a de U est une application linéaire continue $L : E \rightarrow F$ définie lorsqu'elle existe pour tout v comme étant égal à :

$$\lim_{t \rightarrow 0^+} \frac{1}{t} (f(a + tv) - f(a)) = L(v).$$

Avec E et F deux espaces vectoriels normés, et $f : U$ (appartenant à E) $\rightarrow F$

Au sein d'une analyse de sensibilité, ce calcul peut s'effectuer à l'aide de 2 méthodes la procédure d'analyse de sensibilité prospective (FSAP) et la procédure d'analyse de sensibilité ajustée (ASAP). La FSAP est plus facile à développer et à mettre en œuvre que l'ASAP mais elle n'est pas avantageuse lorsqu'il y a plus de variables d'intérêt que de paramètres. Cependant, c'est rarement le cas la plupart des problèmes pratiques étant composés d'un grand nombre de paramètres et de peu de réponses. Après avoir choisi la procédure à utiliser, les sensibilités de tous les paramètres peuvent être calculées, il est alors possible de les utiliser pour classer les différentes variables d'entrée par ordre d'importance, pour évaluer les changements dans la réponse dues aux variations des paramètres ou pour effectuer une analyse d'incertitude.

Cependant il n'est pas nécessaire de plus détailler ces méthodes car contrairement aux méthodes globales qui étudie la variation globale de la sortie $y(z)$ lorsque z varie dans son domaine d'incertitude, les méthodes locales étudient la variation de la sortie $y(z)$ aux alentours de la valeur $z = z_0$. Dans le cadre de cette étude, l'objectif est de déterminer la sensibilité du pH au sein du modèle PaSim, il est pour cela nécessaire de s'intéresser à l'ensemble du domaine de sortie.



Différence entre une analyse de sensibilité locale (gauche) et une analyse de sensibilité globale (droite).
 Graphique tiré de « Analyse d'incertitude, analyse de sensibilité. Objectifs et principales étapes » (2009), par David Makowski, INRA

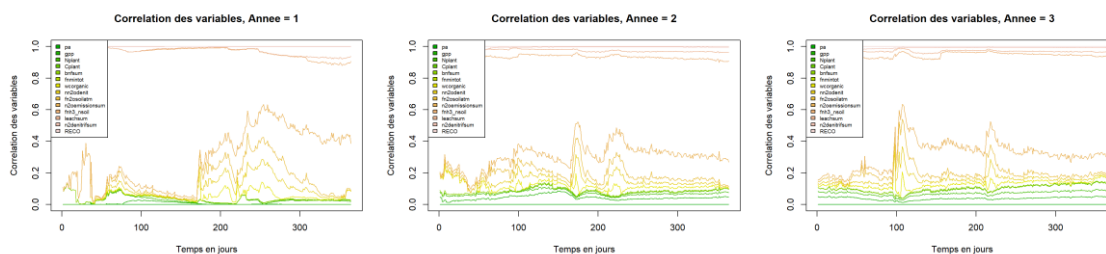
Annexe n°11e : Autre méthode

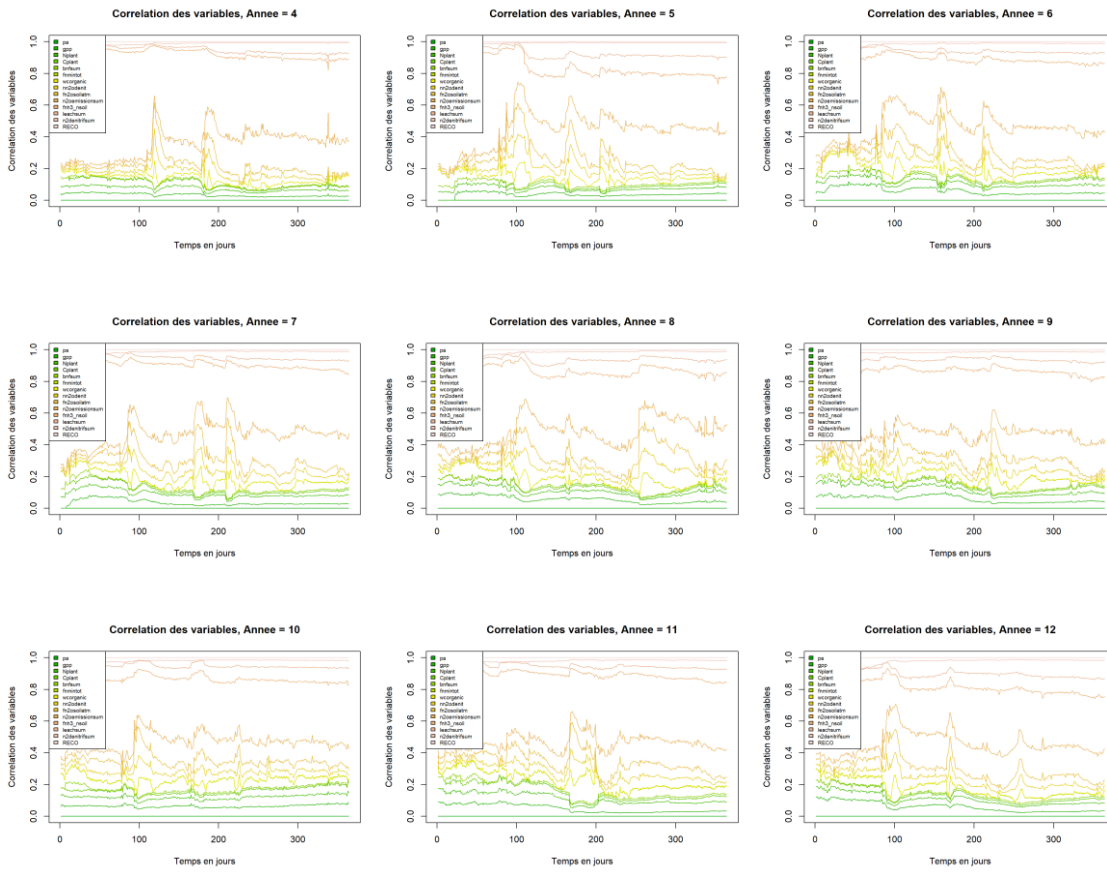
Etude de sensibilité de la VAN (valeur actuelle nette)

En étendant les recherches à d'autres domaines d'étude, il est possible de découvrir que certaines méthodes existantes entre autres en économie peuvent être applicables dans d'autres domaines d'étude. A l'origine, la méthode de sensibilité de la VAN avait pour but de connaître la rentabilité d'un projet.

Cette méthode est proche de la méthode de Morris par l'utilisation d'une valeur d'entrée par défaut et de « saut ». Plus concrètement, pour analyser la sensibilité de la VAN d'un projet, il faut définir les variables d'entrées qu'il est nécessaire d'étudier puis faire varier chacun des paramètres choisis de $\pm 10\%$ et $\pm 15\%$ à partir de sa valeur de base en gardant les autres variables constantes. Il suffit pour finir de présenter les résultats dans un tableau (une table pour chaque variable et un tableau récapitulatif tous les résultats) et sous forme graphique (la valeur de la VAN en fonction du pourcentage de variation).

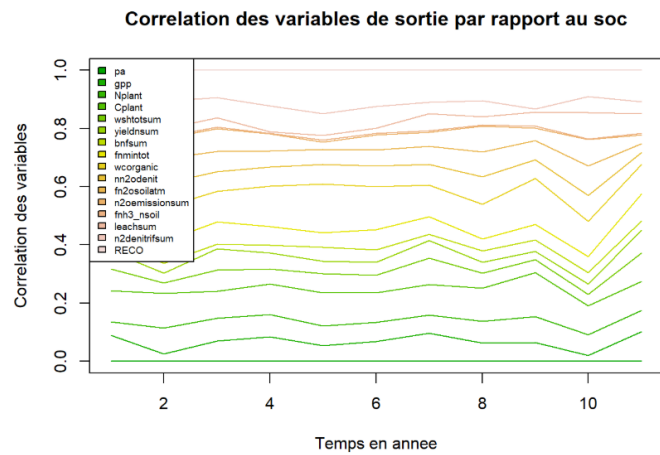
Annexe n°12 : Représentation de Saltelli de la corrélation des variables de sorties avec le pH





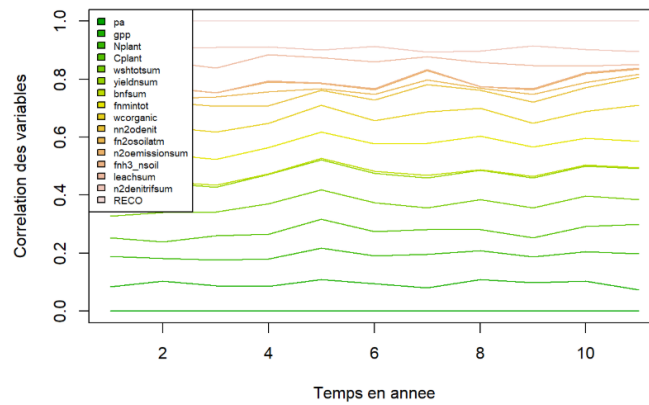
Annexe n°13 : Représentation de Saltelli de la corrélation des variables d'entrée avec les différentes sorties

Variable SOC



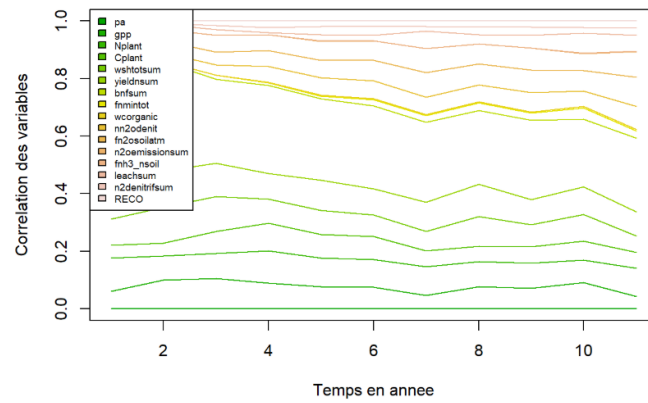
Variable profondeur

Correlation des variables de sortie par rapport a la profondeur



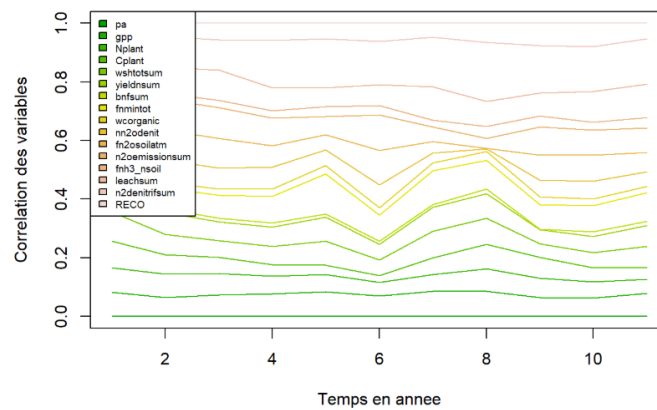
Variable légumineuse

Correlation des variables de sortie par rapport aux legumineuses



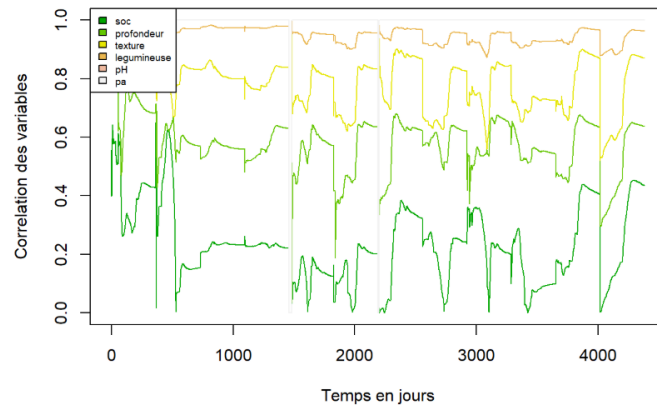
Variable texture

Correlation des variables de sortie par rapport a la texture

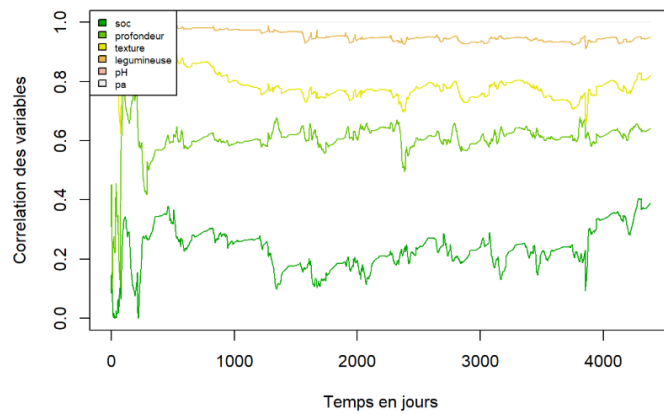


Annexe n°14 : Représentation de Saltelli de la corrélation des variables de sortie avec les différentes entrées

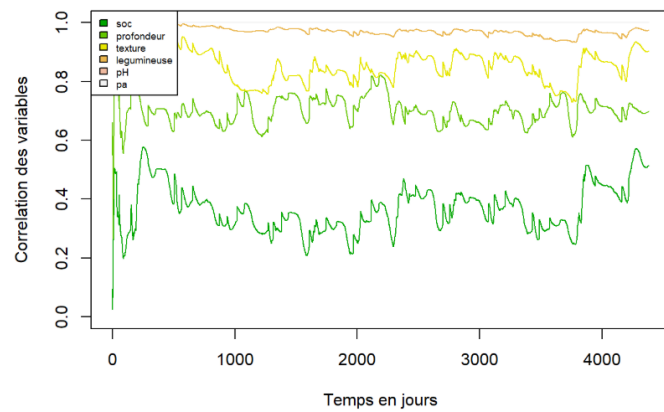
Contribution des entrees sur gpp



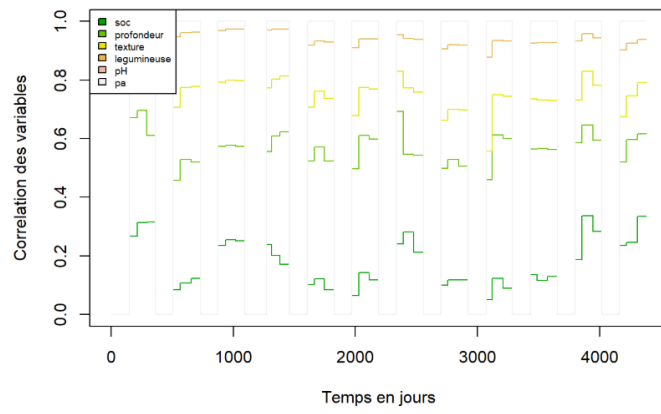
Contribution des entrees sur Nplant



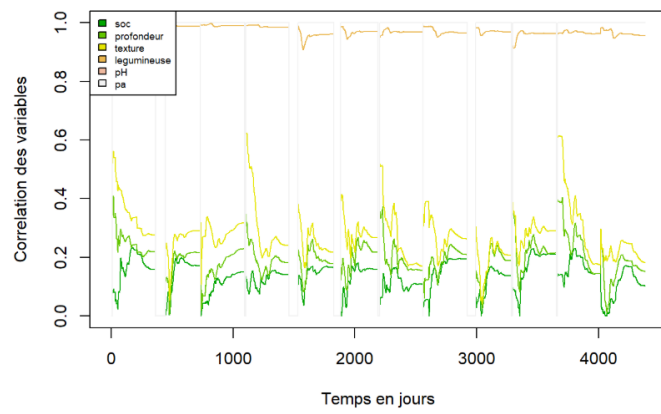
Contribution des entrees sur Cplant



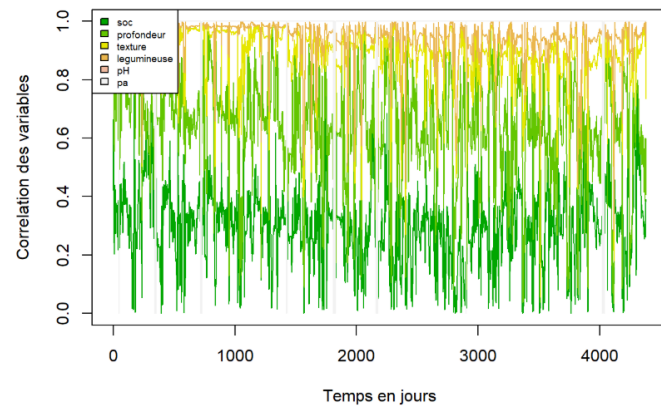
Contribution des entrees sur yieldnsum



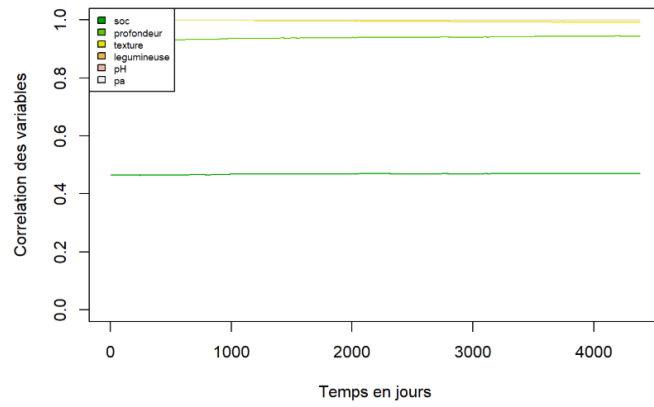
Contribution des entrees sur bnfsum



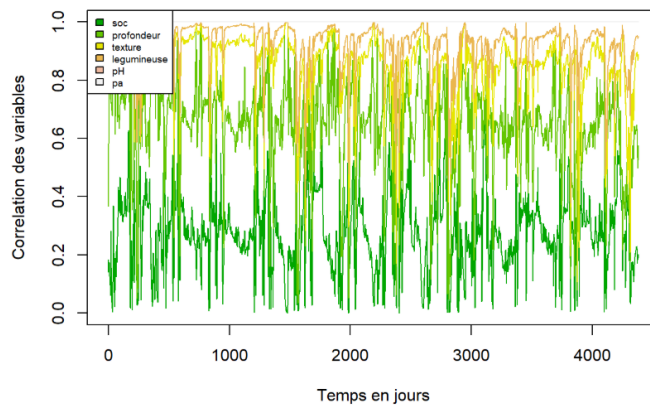
Contribution des entrees sur fmintot



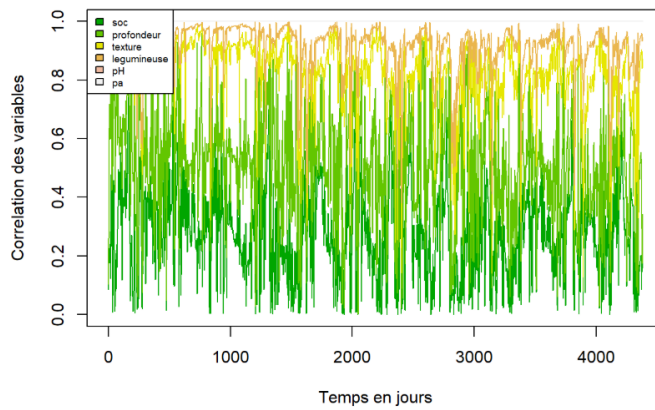
Contribution des entrees sur worganic



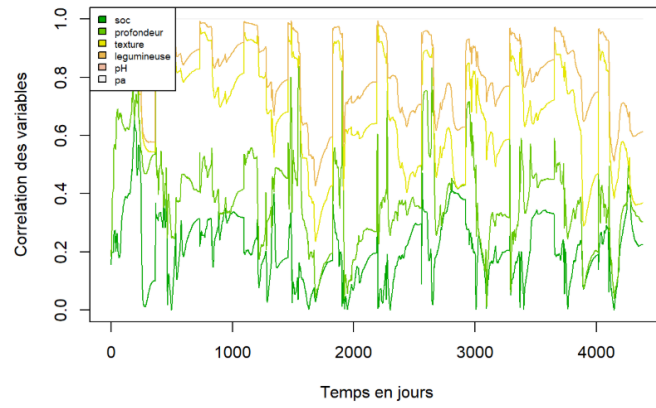
Contribution des entrees sur nn2odenit



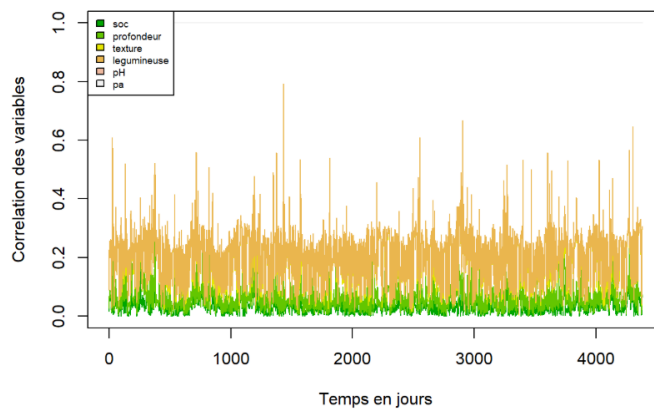
Contribution des entrees sur fn2osoilatm



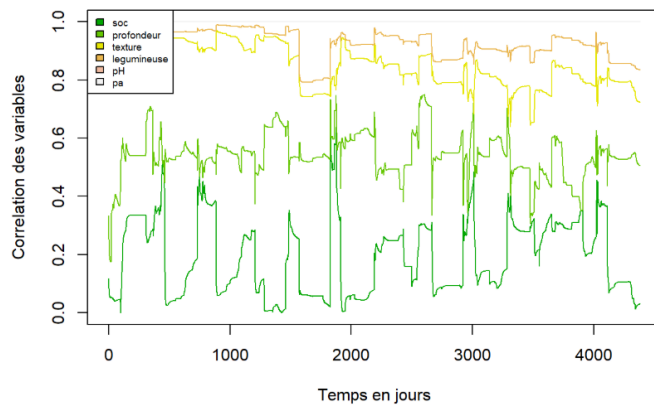
Contribution des entrees sur n2oemissionsum



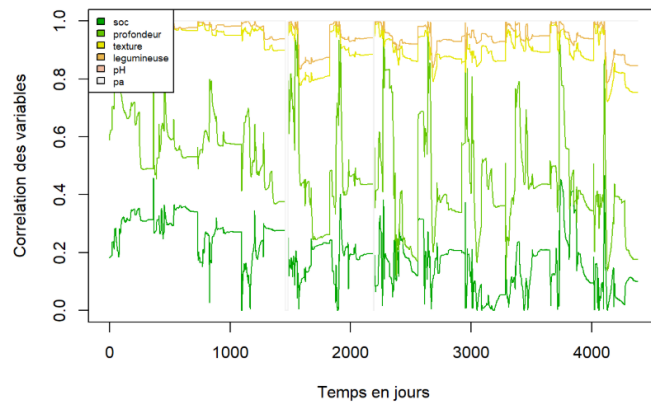
Contribution des entrees sur fnh3_nsoil



Contribution des entrees sur leachsum



Contribution des entrees sur n2denitrifsum



Contribution des entrees sur RECO

