

A D E M E



Agence de l'Environnement
et de la Maîtrise de l'Énergie



Master 1: Statistiques, Probabilités et Mathématiques Appliquées

-Rapport de stage-

*Analyse statistique des erreurs de prédictions de
modèles de simulations des évolutions temporelles des
stocks de carbone dans les sols agricoles*

El-Mehdi BOUNOUA

Période de stage: 15/05/2017 à 15/09/2017

Tuteurs de Stage :

Nicolas Saby, Bassem Dimassi et Manuel Martin (INRA d'Orléans)

Stage réalisé à l'Unité InfoSol de l'INRA d'Orléans :

Avenue de la Pomme de Pin, Ardon BP 20619, 45160 OLIVET Cedex

Remerciements:

Je tiens tout particulièrement à remercier mes maîtres de stage Nicolas Saby, Bassem Dimassi et Manuel Martin pour leur accueil chaleureux et les astuces techniques qu'ils ont pu m'apporter, ainsi que Marion Bardy, la directrice de l'unité InfoSol, de m'avoir accepté en tant que stagiaire au sein de l'INRA d'Orléans.

Ma gratitude va aussi envers Amélia Landre et Mercedes Roman-Dobarco pour leurs conseils pertinents et les codes R qu'elles m'ont fournis.

D'une façon plus générale, je remercie l'équipe d'InfoSol et Sciences du Sol et tous les autres membres de l'équipe avec qui j'ai étroitement collaboré.

Un grand merci tout particulièrement à Bertrand Laroche, Sacha Desbourdes, Stéphanie Guerrier et Eugénie Tientcheu pour leur générosité depuis mon arrivée dans ce service.

Je remercie également Marguerite Zani, professeur encadrant du stage, Laurent Delsol et Didier Chauveau le responsable du master « Statistiques, Probabilités et Mathématiques Appliquées » ainsi que toute l'équipe pédagogique du Collégium Sciences et Techniques d'Orléans pour leurs enseignements m'ayant permis d'effectuer ce stage.

Enfin, je remercie tous les membres de l'INRA qui ont permis mon intégration dans l'Unité d'InfoSol d'Orléans ainsi que pour leur incroyable gentillesse.

Sommaire :

1. Résumé
2. Contexte de stage
 - 2.1. *Présentation générale de l'INRA*
 - 2.2. *L'unité de service InfoSol*
 - 2.3. *Problématique*
 - 2.4. *Objectifs*
 - 2.5. *Démarches*
3. Matériels et méthodes
 - 3.1. *Les données de la base:*
 - 3.1.1. *Les sites et les traitements*
 - 3.1.2. *Domaine géographique*
 - 3.1.3. *Effectif des données*
 - 3.2. *Les types des modèles de simulation :*
 - 3.2.1. *Modèles mécanistes*
 - 3.2.2. *Modèle statistique*
 - 3.3. *Variable à expliquer : Erreur de prédiction sur le taux de changement des stocks de carbone*
 - 3.4. *Les variables*
 - 3.4.1. *Les variables d'essais*
 - 3.4.2. *Les variables de traitements*
 - 3.4.3. *Analyse mono variée de l'erreur*
 - 3.5. *Test de Kruskal-Wallis et régression linéaire simple*
 - 3.6. *Analyse multivariée (ACP)*
 - 3.7. *Packages R*
4. Résultats et discussion
 - 4.1. *Test de Kruskal Wallis*
 - 4.2. *Examen de l'erreur sur le taux de changement des stocks de carbone des modèles en fonction des variables*
 - 4.3. *Analyse multivariée*
5. Conclusion
6. Liste des figures
7. Liste des tableaux
8. Annexe
9. Références

1. Résumé :

Le stockage de carbone organique du sol (COS) est un sujet très important dans les conventions internationales qui visent à lutter contre le changement climatique en réduisant l'effet des émissions des Gaz à Effet de Serre (GES). Des modèles de simulation de ces stocks de carbone ont été conçus pour modéliser la variation temporelle du COS. Ces modèles se basent sur des variables pédoclimatiques et sur des pratiques culturales, on souhaite donc évaluer l'erreur de ces modèles en fonction de ces variables.

Au cours de mon stage, j'ai manipulé à l'aide du logiciel R les données issues de la base de données constituée dans le cadre du projet ITA AMG en effectuant des jointures sur différents tableaux afin de préparer les données en vue d'une étude statistique. J'ai également effectué une analyse statistique en commençant par une approche descriptive des variables liées aux sites et aux pratiques agricoles, ensuite par une approche linéaire de l'erreur sur le taux de changement des stocks de carbone, et enfin une approche multivariée à l'aide de l'analyse en composantes principales. Cette étude a permis aux pédologues de déterminer les facteurs influant l'erreur sur le taux de changement des stocks de carbones dans certains sites et d'hierarchiser les modèles de simulation.

Les principaux résultats sont: Premièrement, l'erreur est faible pour tous les modèles de simulation sauf le modèle ORCHIDEE. Deuxièmement, la corrélation semble très faible avec les propriétés décrivant les sites (Température, pH, teneur d'argile et de calcaire...).

Summary:

The storage of soil organic carbon (SOC) is receiving a great attention from the scientific community because it could help to offset climate change by reducing greenhouse gas (GHG) emission's effect. Process-based models simulating soil organic matter (SOM) dynamics are recognized as valuable tools to quantify and understand SOC dynamics in response to agricultural practices. These models were designed to model the temporal variation of SOC and are based on pedoclimatic variables and on cultural practices, so we aimed to evaluate the error of these models according to these variables. The objective of this work was to analyze the data from carbon simulation models (AMG, Century, ORCHIDEE, RothC ...) and to compare the errors (i.e. carbon stock simulated by the model minus the measured stock) on the rate of change of these stocks.

During my internship, I manipulated data from the database set up under the ITA AMG project, from long-term experiments across France, using software R by performing joins on different tables to prepare data for the statistical analysis. I also carried out a statistical analysis starting with a descriptive approach to variables related to agricultural practices and sites, followed by a linear approach to the error on the rate of change of carbon stocks, finally a multivariate approach using principal component analysis. This study allowed pedologists to determine the factors influencing the error on the rate of change of carbon stocks in some sites and prioritize the simulation models.

The main results were: First, the error is low for all simulation models except the ORCHIDEE model. Secondly, the correlation is very low with the properties describing the sites (temperature, pH, clay and limestone content, etc.).

2. Contexte de stage :

2.1. Présentation générale de l'INRA

Etablissement public à caractère scientifique et technologique (EPST), l'Institut National de Recherche de l'Agronomie (INRA) est placé sous la double tutelle du ministère chargé de la Recherche et du ministère chargé de l'Agriculture. Premier institut de recherche agronomique en Europe, deuxième dans le monde, l'INRA mène des recherches finalisées pour une alimentation saine et de qualité. Créé en 1946, dans un contexte de reconstruction du pays et de modernisation de l'agriculture française, ses objectifs ont évolué et ont désormais atteint une dimension mondiale. En effet, les recherches de l'INRA sont motivées par l'évolution permanente des questionnements scientifiques ainsi que les défis planétaires posés à l'agronomie et l'agriculture tels que le changement climatique, l'épuisement des ressources fossiles et la nutrition humaine.

2.2. L'unité de service InfoSol

Dirigée par Marion Bardy, l'Unité de Service (US) InfoSol fait partie du Département Environnement et Agronomie de l'INRA. La mission principale de cette unité est de constituer et de gérer un système d'information à vocation nationale sur les sols, par rapport à leur distribution spatiale, leurs propriétés et l'évolution de leurs qualités. Elle assure également le contrôle de la qualité de ces données. Elle crée et alimente les bases de données permettant l'exploitation et la traçabilité des informations. Elle contribue à leur diffusion et à leur valorisation en assurant notamment l'articulation avec les bases de connaissances sur les sols et les outils d'exploitation thématique produits par les recherches de l'institut.

L'US InfoSol travaille dans le cadre du groupement d'intérêt scientifique Sol (GIS Sol), sous la tutelle du ministère de l'agriculture et de la pêche, le ministère de l'environnement et du développement durable, l'Institut Géographique National (IGN), l'Institut de Recherche pour le Développement (IRD), et l'Agence de l'Environnement et de la Maîtrise de l'Energie (ADEME). Le GIS SOL propose un ensemble de programmes nationaux pour faciliter et encourager une gestion patrimoniale et durable des sols. Ci-dessous, l'organigramme de l'unité (Figure 1).

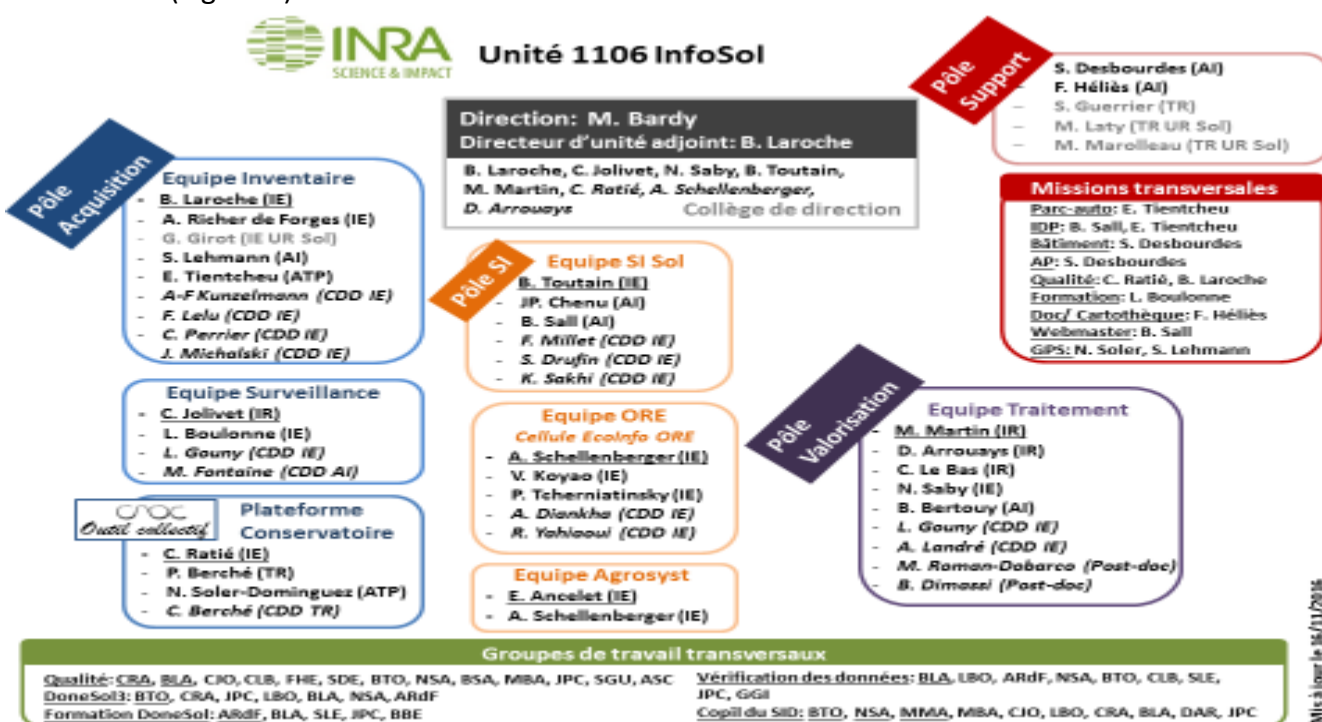


Figure 1: Organigramme de l'unité InfoSol

2.3. Problématique

Le sol est l'épiderme vivant de la terre. Il se définit comme la couche superficielle, meuble, de la croûte terrestre, issue de l'érosion de la roche mère et enrichie par des apports organiques. Cette matière organique qui entre donc dans la composition de tous les sols, leur confère d'importantes propriétés physiques, comme la stabilité structurale (liée à la réduction de la battance) et chimiques comme la capacité à subvenir aux besoins nutritionnels des plantes (en libérant des éléments minéraux comme l'azote, le soufre ou le phosphore) (Lal et al., 2007; Zhang et al., 2014)

Les matières organiques du sol assurent également de nombreuses fonctions agronomiques et environnementales. Elles améliorent la fertilité, l'aération, la réserve en eau et la biodiversité du sol. Elles limitent la compaction et l'érosion hydrique et favorisent le piégeage des métaux toxiques ou des micropolluants organiques. Elles proviennent de la transformation des débris végétaux par les organismes vivants. Composées de 58 % de carbone organique en moyenne, elles libèrent du dioxyde de carbone (CO₂) et des composés organiques en se décomposant sous l'influence du climat et des conditions ambiantes du sol (Smith et al. 2012). Ce gaz participe activement au cycle des Gaz à Effet de Serre (GES) ; en absorbant le rayonnement infrarouge émis par la surface terrestre et qui conduit au réchauffement climatique.

Cette préoccupation d'ordre environnemental, est mise en avant depuis quelques années. Lorsque la matière organique du sol est dégradée par la biomasse microbienne, du dioxyde de carbone est relâché dans l'atmosphère. Le protocole de Kyoto (un accord international visant à la réduction des émissions de gaz à effet de serre) a souligné l'importance du sol dans les émissions de CO₂. A l'heure de la lutte contre le réchauffement climatique, le sol peut jouer un rôle important dans les deux sens, aussi bien en tant que source d'émissions qu'en tant que puit, en stockant le carbone piégé par les plantes sous forme de matière organique (IPCC 2014).

Les engagements internationaux visant à lutter contre ce phénomène prennent en compte, depuis plusieurs années, la variation des stocks de carbone organique dans le sol (COS). La Convention Cadre des Nations Unies sur le Changement Climatique (CCNUCC, signée en 1992) et le Protocole de Kyoto (signé en 1997) ont introduit, pour les pays signataires, la réalisation annuelle d'un inventaire des émissions et des absorptions de gaz à effet de serre. Dès lors, il devient primordial de prendre en compte avec exactitude la variation des stocks du carbone organique du sol dans le cadre de ces négociations internationales. De plus, la connaissance des phénomènes influents sur la variation des stocks de COS permet, à des échelles locales, de mettre en place des pratiques plus adaptées visant à accroître les stocks. (IPCC 2014).

Ainsi, en collaboration avec l'ADEME, INRA, Arvalis, Citepa, Agro-Transfert RT et LDAR, le projet Csopra dans lequel s'inscrit mon sujet de stage, a pour but de développer des méthodes de comptabilisation du stockage de carbone organique des sols sous l'effet des pratiques culturales et d'améliorer la méthodologie d'inventaire.

Certes en réalisant des mesures tous les ans, il est possible de voir l'évolution du carbone dans nos sols de manière précise. Mais, ces mesures sont onéreuses (surtout quand il s'agit d'établir une étude au niveau national) et ne permettent pas de prédire l'évolution future des sols, d'où l'utilité d'utiliser des modèles de simulation afin d'évaluer à long terme le stock du COS sous l'effet de différents facteurs pédoclimatiques et les pratiques culturales.

A l'aide de ces modèles, qui s'appuient sur des variables d'entrées qui diffèrent d'un modèle à un autre, et grâce à quelques mesures faites sur des sites répartis le long de la France métropolitaine nous pouvons évaluer la précision et l'erreur de certains modèles mécanistes

(AMG (Andriulo et al. 1999), CENTURY (Parton et al. 1988), ORCHIDEE (Kinner et al. 2005), RothC) et un modèle statistique (IPCC Tier1 (IPCC 2014)) afin de choisir celui qui s'adapte le mieux aux sols de France, et qui prend en compte les pratiques culturales.

2.4. Objectifs :

Mon stage comporte deux objectifs:

-Déterminer les facteurs influant l'erreur sur le taux de changement des stocks de carbone des modèles de simulation : Nous souhaitons déterminer la distribution des erreurs de prédiction des modèles mécanistes et statistiques de simulation.

-Hiérarchiser les modèles de prédiction : Un classement des modèles de simulation en terme d'erreur de prédiction à l'aide d'une analyse statistique de l'erreur sur le taux de changement des stocks de carbone dans les sols.

2.5. Démarches :

Pour atteindre ces objectifs, une analyse descriptive de l'erreur sur le taux de changement des stocks de carbone sera établie ainsi que l'analyse des variables pédoclimatiques et des pratiques culturales. Ensuite un test de Kruskal-Wallis pour comparer les modèles de simulation et évaluer la précision de chaque modèle. Après, une régression linéaire simple afin de voir l'effet de chaque variable sur l'erreur. Et enfin, une analyse multivariée: l'analyse en composante principale pour mettre en évidence les corrélations entre l'erreur des modèles et les variables liées aux sites.

3. **Matériels et méthodes :**

3.1. **Les données de la base:**

3.1.1. **Les sites et les traitements.**

Les données du carbone et des variables pédoclimatiques sont issues de la base de données constituée dans le cadre du projet ITA AMG à partir d'essais de durées moyennes à longues (de 8 à 40 ans). Elles ont été mesurées au cours d'essais expérimentaux sur différents sites de la France. Les essais sont subdivisés en traitements. Dans chaque site, l'objectif est de tester l'effet d'un ensemble de pratiques agricoles. De plus, un site se caractérise par un lieu et une période de temps. Chaque traitement d'un site correspond à un type de pratiques agricoles testé. En outre, les sites sont découpés en blocs, des parcelles élémentaires adjacentes, de sorte à avoir des caractéristiques pédologiques les plus homogènes possibles au sein de chaque bloc.

Le nombre de traitements dans un site varie d'un site à l'autre.

3.1.2. **Domaine géographique**

Les sites de la base de données sont répartis sur la France métropolitaine, un domaine sur lequel on souhaite évaluer la prédiction des modèles de simulation. (Figure 2)



Figure 2: carte géographique des sites d'essais

3.1.3. Effectif des données

Tableau 1: Tableau des effectifs de la base de donnée

Nombre d'essais	Nombre de traitements	Nombre d'observations
19	89	614

3.2. Les types des modèles de simulation de carbone:

3.2.1. Modèle mécaniste

Un modèle mécaniste est une représentation mathématique d'un ou plusieurs processus caractérisant le fonctionnement d'un système bien délimité qui peut être biologique, économique, etc. Généralement ces modèles consistent en un ensemble d'équations différentielles qui définissent chaque processus comme étant une fonction de lois physiques ou des connaissances empiriques. Les sorties d'un processus peuvent être les entrées d'un autre (Dubitzky et al., 2013). Dans notre cas les modèles utilisés sont des modèles biogéochimiques qui simulent la dynamique du C du sol et plusieurs autres processus tel que la croissance et développement des cultures, la dynamique de l'azote, etc.

3.2.2. Modèle statistique

Un modèle statistique est une description mathématique du mécanisme qui a généré les observations, que l'on suppose être un processus stochastique et non un processus déterministe. Il s'exprime généralement à l'aide d'une famille de distributions et d'hypothèses sur les variables aléatoires. Les principaux types sont : le modèle linéaire, le modèle linéaire généralisé, le modèle mixte, etc. (Davidson, 2008).

3.3. Variable à expliquer : Erreur sur le taux de changement des stocks de carbone des modèles de simulation

Les stocks de carbone organique du sol sont suivis parcelle élémentaire par parcelle élémentaire. De par les conditions de collecte des jeux de données, les modes de suivi de l'évolution dans le temps des stocks de carbone varient d'un essai à l'autre, voire d'un traitement à l'autre, et parfois au cours du temps, au sein d'un même essai.

Le stock de carbone n'est en réalité pas mesuré directement. On mesure en fait en laboratoire une teneur en carbone sur un échantillon prélevé pour être représentatif d'une certaine couche de sol de la parcelle élémentaire - Seule la première couche est considérée dans cette étude. On calcule ensuite la masse de cette couche par hectare grâce à la densité de sa terre fine et de sa teneur en élément grossiers. En multipliant la masse de terre fine par sa teneur en carbone, on obtient un stock de carbone par hectare.

Les données obtenues dans un même traitement et à une même date sont ensuite moyennées, car nous souhaitons nous affranchir au maximum des erreurs de mesure et de la variabilité spatiale du stock de carbone, et le traitement est la plus grande unité au sein de laquelle les facteurs influant sur la dynamique du carbone du sol sont théoriquement homogènes. Dans ce cas nous faisons l'hypothèse que les mesures des parcelles élémentaires sont issues de la même population et indépendantes. Ensuite, l'évolution du stock de carbone de chaque traitement au cours du temps est simulée par les différents modèles.

La variable que nous cherchons à expliquer est l'erreur de prédiction sur le taux de changement des stocks de carbone qui a été calculée à plusieurs dates pour chaque traitement

(à chaque fois qu'une mesure de stock de carbone est disponible) : nous disposons donc de mesures répétées au cours du temps de la même variable. Ci-dessous (Formule 1 et 2), la définition de l'erreur de prédiction et l'erreur sur le taux de changement des stocks de carbone dans les sols.

Formule de l'erreur de prédiction : $E_C = C_{Sim} - C_{Mes}$ (1)

Avec C_{Sim} : Stock de carbone simulé ; C_{Mes} : Stock de carbone mesuré

Formule de l'erreur sur le taux de changement des stocks: $R_{err} = R_{Sim} - R_{Mes}$

Avec $R_{Sim} = \frac{C_{Sim}(t_f) - C_{Sim}(t_0)}{t_f - t_0}$; $R_{Mes} = \frac{C_{Mes}(t_f) - C_{Mes}(t_0)}{t_f - t_0}$

Où t_0 la date du début de l'essai et t_f la date finale en année

L'erreur sur le taux de changement des stocks de carbone prend en compte la durée de l'essai et modélise donc la pente d'évolution de l'erreur de prédiction pour chaque traitement.

3.4. Les variables :

A priori, les erreurs de prédiction des modèles de simulation de carbone dépendent des variables qui caractérisent les sites (variables abiotiques) et les traitements (pratiques culturales).

3.4.1. Les variables d'essais : (abiotiques)

Les variables liées aux sites sont d'abord de nature pédoclimatique : température, différence entre précipitations et évapotranspirations annuels (P-ETP), teneur du sol en argile et en calcaire, le pH du sol et le stock initial de carbone.

Le tableau 2 ci-dessous résume les statistiques descriptives de ces variables sur les 614 observations.

Tableau 2: Statistiques descriptives des variables abiotiques

	unité	minimum	maximum	Médian	Moyenne
Température	°C	9,89	13,43	11,19	11,45
P-ETP	mm/an	-272,17	469,30	-59,81	-11,30
Argile	g.kg ⁻¹	43,00	354,00	196,50	199,61
CaCO ₃	g.kg ⁻¹	0,00	781,00	0,00	62,22
pH	sans unité	5,85	8,33	7,28	7,25
Stock initial de carbone	t.ha ⁻¹	25.08	77.71	42.42	45.49

Le calcul du kurtosis (coefficient d'aplatissement) et du skewness (coefficient de dissymétrie) pour chaque variable. Les résultats dans le tableau 3

Kurtosis : c'est une mesure de l'aplatissement de la distribution d'une variable. Pour une variable X de moyenne μ et d'écart-type σ , on définit son kurtosis non normalisé comme le moment d'ordre 4 de la variable centrée réduite (sous réserve d'existence) (formule 2) :

$$K = \left[\left(\frac{X - \mu}{\sigma} \right)^4 \right] \quad (2)$$

Skewness : c'est une mesure de l'asymétrie de la distribution d'une variable. Pour X, On définit le skewness comme le moment d'ordre 3 de la variable centrée réduite (sous réserve d'existence) (formule 3):
$$S = E \left[\left(\frac{X - \mu}{\sigma} \right)^3 \right] \quad (3)$$

Tableau 3: kurtosis et skewness des variables abiotiques

	kurtosis	skewness
Température	-0,20	0,67
P-ETP	-0,14	0,81
Argile	-0,48	0,33
CaCO3	17,19	3,93
pH	-1,43	-0,14
Stock initial de carbone	0.90	0.92

On trace les histogrammes de chaque variable (les traitements de tous les sites confondus) afin de visualiser la distribution :

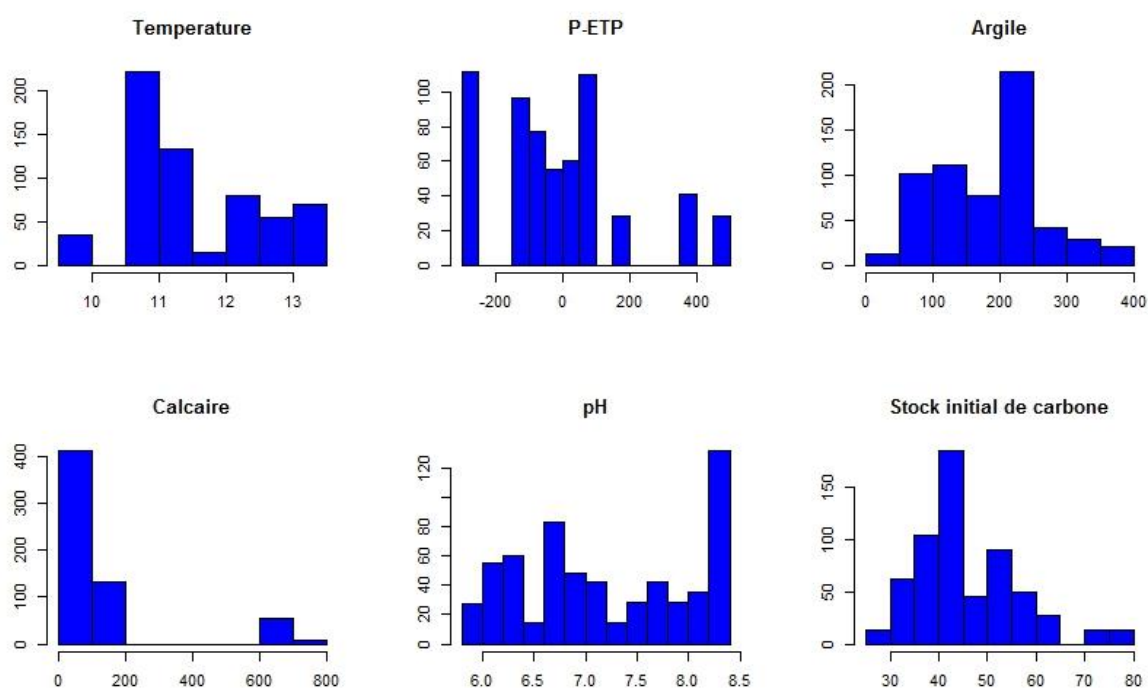


Figure 3: Histogrammes: distribution des variables abiotiques

3.4.2. Les variables de traitement :

Les variables de traitement sont liées aux pratiques, on retrouve le travail du sol (labour), les cultures intermédiaires, les résidus végétaux et animaux, l'ajout de matière organique au début de l'essai et la culture de Colza.

Un codage numérique est adapté pour l'analyse des quatre pratiques :

- **Fertilisants Organique** : apport de fertilisant organique (0: minimum; 1 : ajout max) (Exemple : fumier).
- **Travail du sol** : labour (0 : minimum; 1.5: labour max).
- **Résidus de culture** : résidus de culture laissés au champ (0: pas de résidus ;2: résidus max).
- **Cultures intermédiaires** : cultures intermédiaires (0 pas de cultures intermédiaires, 1 : cultures intermédiaires tous les ans).
- **Culture Colza** : culture de colza (0 pas de culture, 1 culture tous les ans).

Une analyse mono-variée des pratiques est représentée dans le tableau 4 et 5 ci-dessous :

Tableau 4: Statistiques descriptives des pratiques culturales

	Minimum	Maximum	Médian	Moyenne
Fertilisants Organiques	0	1	0.94	0.86
Travail du sol	0.19	1.34	0.95	0.75
Résidus	0	0.35	0	0.03
Culture intermédiaire	0	0.39	0	0.026
Culture colza	0	0.45	0	0.052

Tableau 5: Kurtosis et skewness des pratiques culturales

	kurtosis	skewness
fertOrg	7,16	-2,90
Travail du sol	-1,19	-0,55
résidus	10,37	3,09
culture intermédiaire	3,99	-1,04
Culture de colza	5.16	2.44

Les histogrammes de pratiques agricoles :

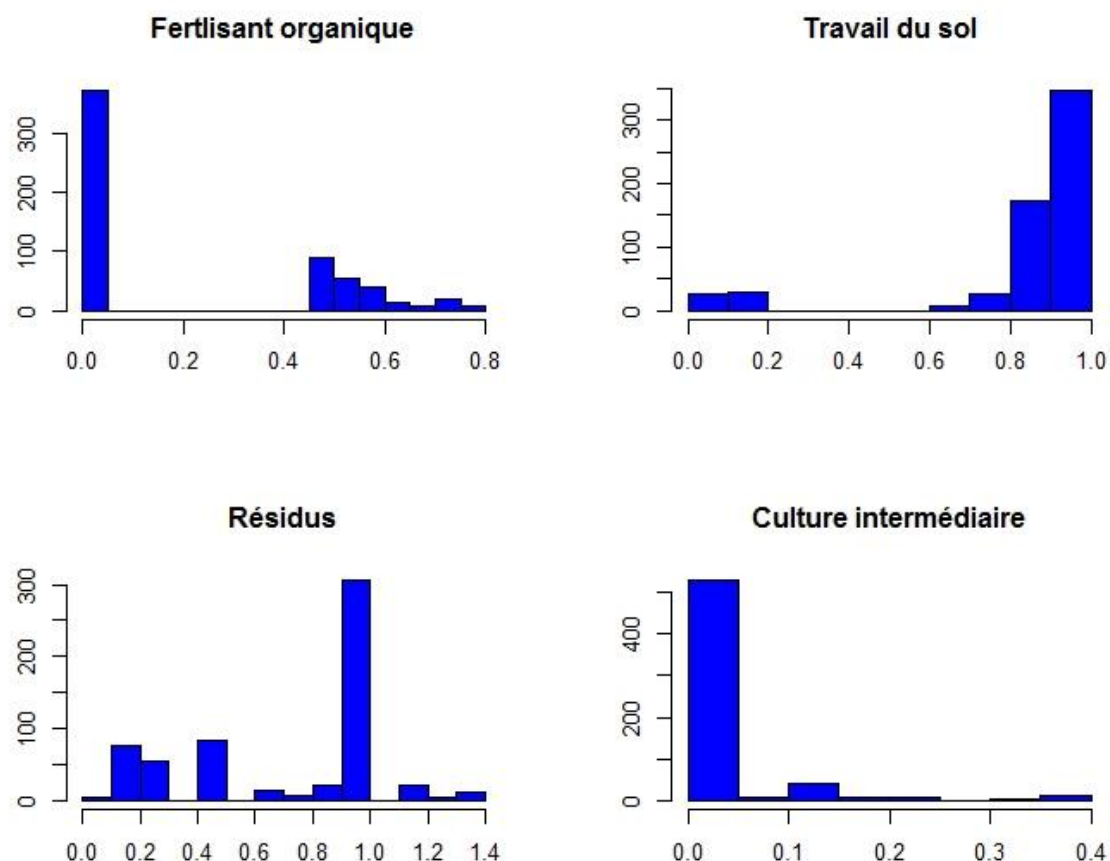


Figure 4: Histogrammes: distribution des pratiques agricoles

3.4.3. Analyse descriptive de l'erreur sur le taux de changement des stocks de carbone :

Une étude semblable à l'analyse des variables est effectuée sur l'erreur sur le taux de changement des stocks de carbone, tout modèle compris. Les statistiques descriptives ainsi que l'histogramme de la répartition sont représentés dans le tableau 6 et la figure 7 ci-dessous.

Tableau 6: statistiques descriptives de l'erreur de prédiction

	Min	Médian	Moyenne	Max	Kurtosis	Skewness
Erreur globale sur le taux de changement des stocks de carbone (t.ha-1.an-1)	-4.19	0.10	0.12	4.15	13.68	-0.68

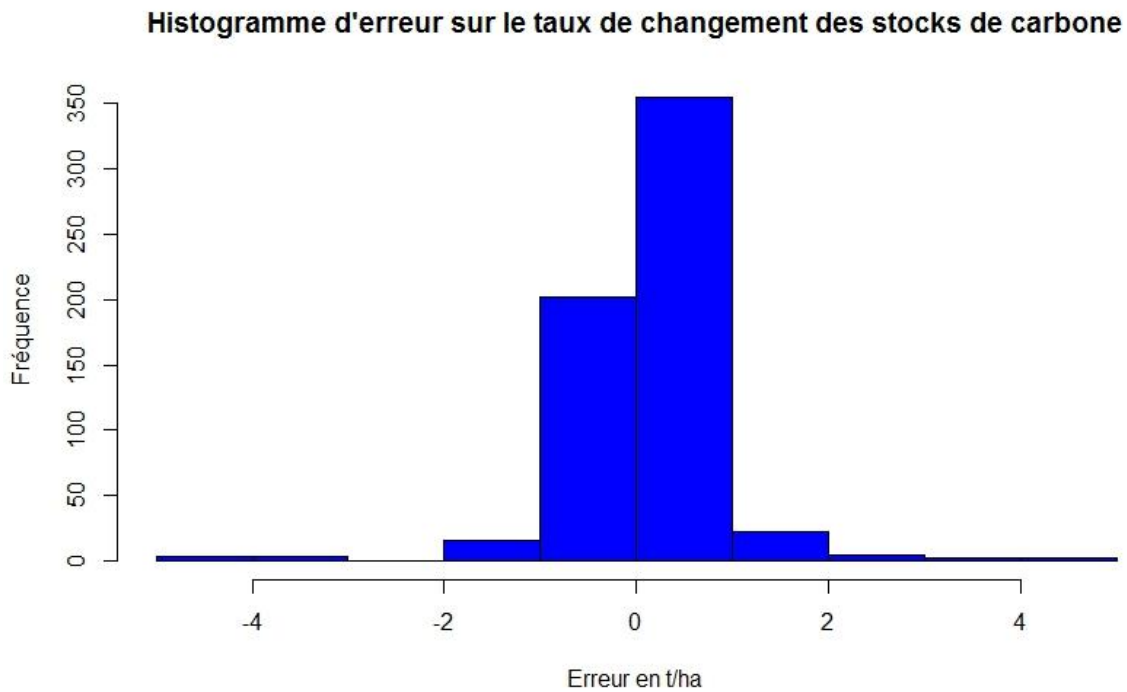


Figure 5: Histogramme: distribution de l'erreur sur le taux de changement des stocks de carbone

3.5. Test de Kruskal-Wallis et régression linéaire simple

L'application du test de Kruskal-Wallis sur l'erreur sur le taux de changement des stocks de carbone pour chaque modèle permet de comparer globalement les modèles de simulation. Ce test non paramétrique va donc révéler si les modèles sont semblables ou différents grâce aux lettres (de "a" à "e") affichées au-dessus de chaque box plot (Figure 6). C'est une alternative de l'ANOVA (Analyse de la variance) qui, ici, ne peut être appliquée par manque d'hypothèse d'homoscédasticité. La deuxième approche est de tracer le taux de stock de carbone simulé en fonction du taux de stock mesuré pour voir la répartition du nuage de points. Un modèle est plus précis si son nuage de point est « regroupé » autour du premier axe du repère ($y=x$).

Test de Kruskal-Wallis :

Principe du test:

Si on désigne par M_i le paramètre des modèles, les hypothèses nulle H_0 et alternative H_a du test de Kruskal-Wallis sont les suivantes :

$$H_0: M_1 = M_2 = \dots = M_k$$

$$H_a: \text{il existe au moins un couple } (i, j) \text{ tel que } M_i \neq M_j$$

Le calcul de la statistique K du test de Kruskal-Wallis fait intervenir le rang des observations, une fois les k échantillons (ou modèles) mélangés. K est défini par (formule 3):

$$K = \frac{12}{N(N+1)} \sum_{i=1}^k (n_i * R_i^2 - 3(N+1)) \quad (3)$$

Où n_i est la taille de l'échantillon i , N la somme des n_i , et R_i la somme des rangs pour l'échantillon i parmi l'ensemble des échantillons.

Une fois la précision des modèles testée, on souhaite repérer les variables pour lesquelles l'erreur sur le taux de changement de stock de carbone est sensible. On trace cette erreur en fonction de chaque variable et on fait un test d'hypothèse sur la droite de régression linéaire simple. Les points du nuage correspondent à des moyennes sur les années pour chaque traitement. Sur R, à l'aide du Package (ggplot2) on trace l'évolution de ce taux regroupée par modèle de simulation, la droite de régression ainsi que l'intervalle de confiance au niveau 95% autour de cette droite. Un tableau effectuant le test d'hypothèse sur la pente et l'ordonnée à l'origine des droites de régression et le coefficient de détermination (R-carré) qui est le rapport entre la variance expliquée et la variance totale (formule 4):

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

Principe du modèle linéaire et du test sur la droite de régression linéaire simple :

Notons l'erreur sur le taux de changement des stocks de carbone "X" qui est un phénomène supposé aléatoire (dépendant de la simulation des modèles). Ce phénomène est influencé par les p variables abiotiques et les pratiques culturales R_i avec $i \in \{1, 2, \dots, p\}$ (aussi appelé les régresseurs). On suppose une relation linéaire (ou affine) entre X et les R_i . On suppose aussi observer la somme de l'effet de ces régresseurs et d'un ensemble de perturbations non observables, que l'on résume par un bruit gaussien centré. Ce modèle s'écrit ainsi (Formule 5):

$$X = \beta + \sum_{j=1}^p \alpha_j R_i^j + \varepsilon_i \quad (5) \quad \text{avec } \varepsilon_i \sim \mathcal{N}(0, \sigma^2)$$

Ici on analyse l'erreur en fonction d'une variable. (p=1)

L'équation devient (Formule 6) :

$$X_i = \beta + \alpha R_i + \varepsilon_i \quad (6)$$

Avec $i \in \{1, 2, \dots, n\}$ et n est le nombre d'observation.

Le calcul des estimateurs donne (Formule 7 et 8):

$$\hat{\alpha} = \frac{Cov(R, X)}{\sigma_R^2} \quad (7)$$

$$\hat{\beta} = \bar{X} - \hat{\alpha} \bar{R} \quad (8)$$

Avec \bar{R} : La moyenne empirique de R

\bar{X} : La moyenne empirique de X

σ_R^2 : La variance empirique de R

On en déduit l'expression des variances ainsi que la covariance entre les deux estimateurs ($\hat{\alpha}$ et $\hat{\beta}$ ne sont pas indépendants). Comme ils sont des estimateurs sans biais des paramètres qu'ils estiment et suivant des lois normales (combinaison linéaire de X), on a :

$$\hat{\alpha} \sim \mathcal{N}\left(\alpha, \frac{\sigma^2}{n\sigma_R^2}\right)$$

$$\hat{\beta} \sim \mathcal{N}\left(\beta, \frac{\sigma^2 \sum_{i=1}^n R_i^2}{n\sigma_R^2}\right)$$

La connaissance des lois des estimateurs de (β, α) qui dépendent de σ^2 ainsi que de la loi de l'estimateur de σ^2 et la propriété d'indépendance permet de construire des intervalles de confiance autour de la droite de régression et des tests pour α et β .

L'estimateur Z de α centré réduit et de loi normale et indépendant de l'estimateur de σ^2 normalisé

$$Z = \frac{(\hat{\alpha} - \alpha)\sqrt{n\sigma_R}}{\hat{\sigma}} \sim \mathcal{N}(0,1) , \quad \frac{(n-p-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-p-1)$$

D'où la loi de Student (Formule 8):

$$\frac{(\hat{\alpha} - \alpha)\sqrt{n\sigma_R}}{\hat{\sigma}} = \frac{(\hat{\alpha} - \alpha)}{\hat{\sigma}(\hat{\alpha})} \sim t(n-p-1) \quad (8)$$

Où $\hat{\sigma}(\hat{\alpha}) = \hat{\sigma}/\sqrt{n\sigma_R}$ est l'estimateur de l'écart-type de l'estimateur de α . On en déduit un intervalle de confiance de Student de niveau $(1-\gamma)$ pour α de la forme $\hat{\alpha} \mp \hat{\sigma}(\hat{\alpha})t_{(n-p-1, 1-\gamma/2)}$

Et le test de Student pour $H_0 : \alpha=0$ qui a pour statistique de test (Formule 9)

$$T = \frac{\hat{\alpha}}{\hat{\sigma}(\hat{\alpha})} \sim t(n-p-1) \quad (9) \quad \text{sous } H_0$$

3.6. Analyse multivariée :

Une autre méthode pour analyser la corrélation entre l'erreur sur le taux de changement des stocks de carbones des modèles, les variables abiotiques et les pratiques culturelles, est l'analyse en composante principale sur l'ensemble des variables quantitatives.

Le but de cette méthode est de projeter l'ensemble du nuage de points préalablement représenté en dimension p (le nombre des variables quantitatives) sur un plan de dimension 2 ou chaque axe contient un maximum d'informations, pouvant être confondus, sur l'ensemble des variables. Un pourcentage d'inertie est donné pour évaluer la perte d'information due au passage d'une grande dimension à la représentation planaire.

3.7. Packages R

De la préparation de données à partir de tableaux Excel, à la représentation des graphes, nous aurons besoin d'un grand nombre de Packages R afin de bien cerner l'analyse et bien organiser les scripts R. Ci-dessous, tous les Packages utilisés et leurs fonctions :

-tidyr : Pour utiliser les fonction « gather » et « spread » pour manipuler les colonnes en fonction d'une autre colonne dans un d-frame (amélioré de Reshape)

-plyr : C'est un ensemble d'outils propres et cohérents qui met en application le modèle split-apply-combine sur R.

-dplyr : Très utile pour la manipulation de données. Egalement utilisé pour le pipelines (communication interprocessus)

-ggplot2 : Outil idéal pour tracer des graphes élégants, des boxplots et des histogrammes.

-factoextra : Permet de tracer les graphes d'ACP, utilise en entrée une « PCA » de FactoMineR

-FactoMineR : Pour utiliser la fonction PCA qui effectue l'analyse en composante principale

-e1071 : Pour les fonction kurtosis et skewness

- **soiltexture** : pour tracer le triangle de texture des sols
- **MASS**: pour l'utilisation de la fonction qui effectue le test de Kruskal-Wallis
- **multcompView**: Convertie le vecteur logique d'identification du test de Kruskal-Wallis en affichage.

4. Résultats et discussion:

4.1. Test de Kruskal Wallis

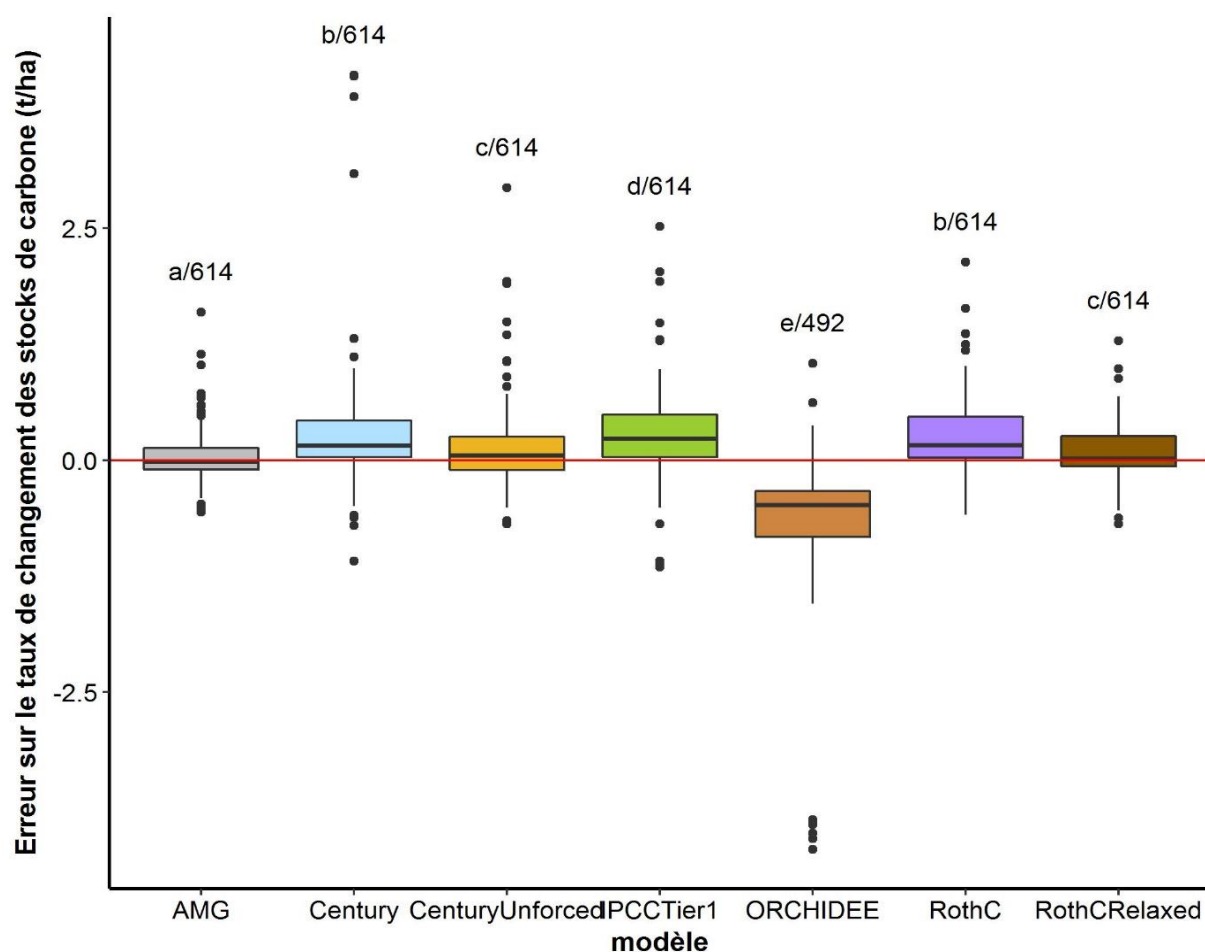


Figure 6: Boxplots des erreurs de prédiction par modèle de simulation. Deux boxplots avec la même lettre sont identiques d'après le test. Le nombre correspond au nombre d'observation pour chaque modèle.

D'après les Box plots (figure 6) les modèles se divisent en quatre groupes: le groupe du modèle AMG (lettre "a") qui est le plus précis avec une répartition quasi symétrique de l'erreur. Le groupe des modèles RothC et Century viennent en 2^{ème} position avec une légère sous-estimation de l'erreur (lettre "b"). Ensuite, RothC Relaxed et Century Unforced dans le troisième groupe qui surestiment l'erreur (lettre "c") suivi du modèle IPCC Tier1 qui sous-estime l'erreur moyenne et enfin ORCHIDEE qui a une erreur considérable. Ci-dessous (Tableau 7) les moyennes d'erreurs ainsi que les écarts-type suivant chaque modèle de simulation.

Tableau 7: Tableau des moyennes et des écarts-types de l'erreur par modèle de simulation

Modèle	AMG	RothC Relaxed	Century	Century Unforced	IPCC Tier1	ORCHIDEE	RothC
Moyenne	0.09	0.11	0.43	0.25	0.34	-0.79	0.32
Ecart-type	0.37	0.35	0.90	0.57	0.64	1.06	0.47

La figure 7 représente le taux de carbone simulé en fonction du mesuré. Comme on a 19 sites et pour un souci de clarté, on choisit une couleur et un symbole pour représenter un site. La droite en noire est le premier axe du repère ($y=x$).

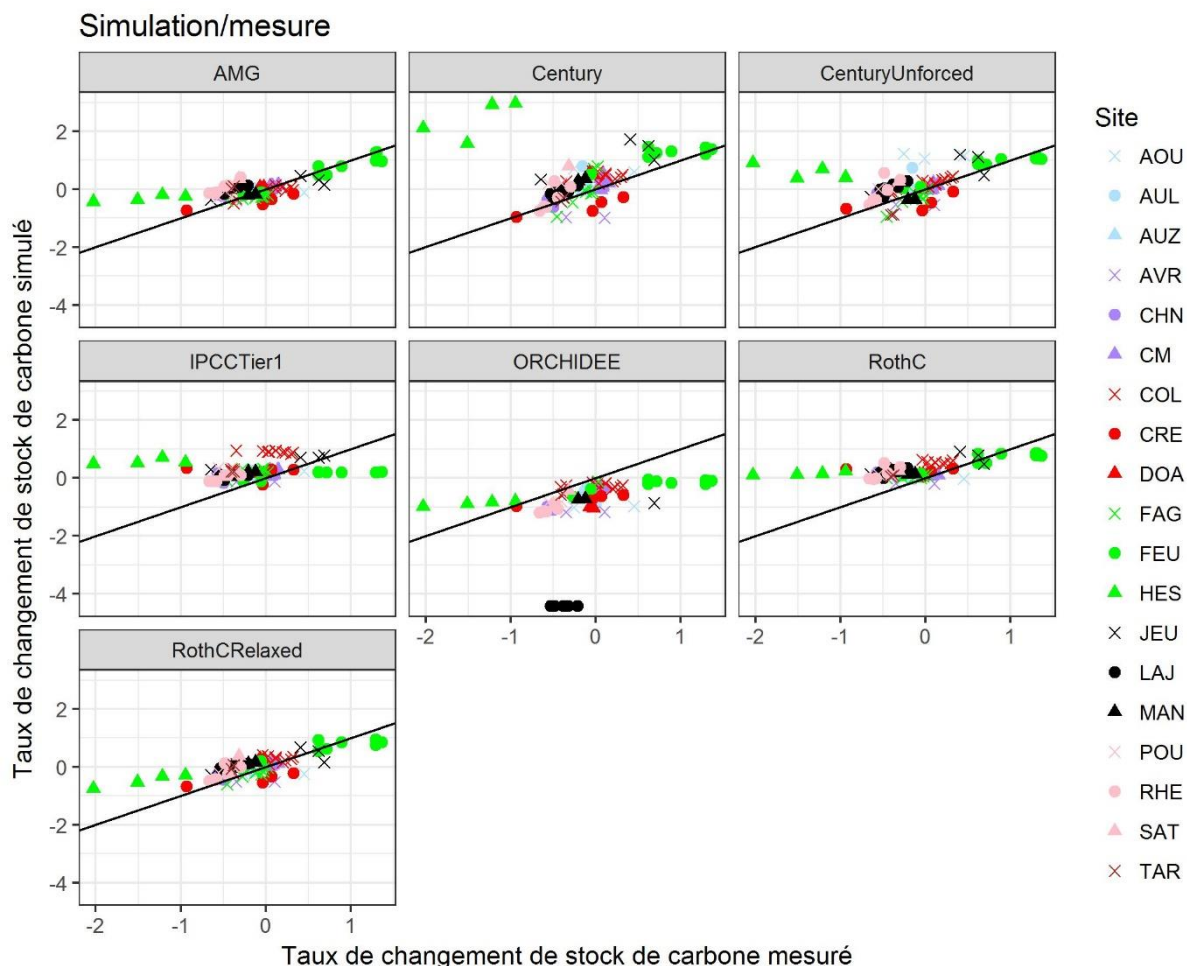


Figure 7: Nuage de points du taux de stockage de carbone simulé en fonction du taux de stockage de carbone mesuré. La droite $y=x$.

Nous retrouvons encore une fois pour les modèles précis (AMG, RothC Relaxed) un regroupement du nuage autour de l'axe $y=x$ et des différences considérables pour ORCHIDEE et RothC. Les valeurs aberrantes de l'erreur sont dans les sites Hessange et Feucherolles.

4.2. Examen de l'erreur sur le taux de changement des stocks de carbone des modèles en fonction des variables

La distribution des erreurs de prédiction des modèles par rapport aux variables abiotiques et aux pratiques culturales ainsi qu'un tableau regroupant les tests sur les pentes de régression et l'ordonnée à l'origine des modèles est présentée ci-dessous (de la Figure 8 à la Figure 18 et tableau 8 à 18). Les cases rouges pour les p-valeur $< 5\%$ (test significatif):

Température :

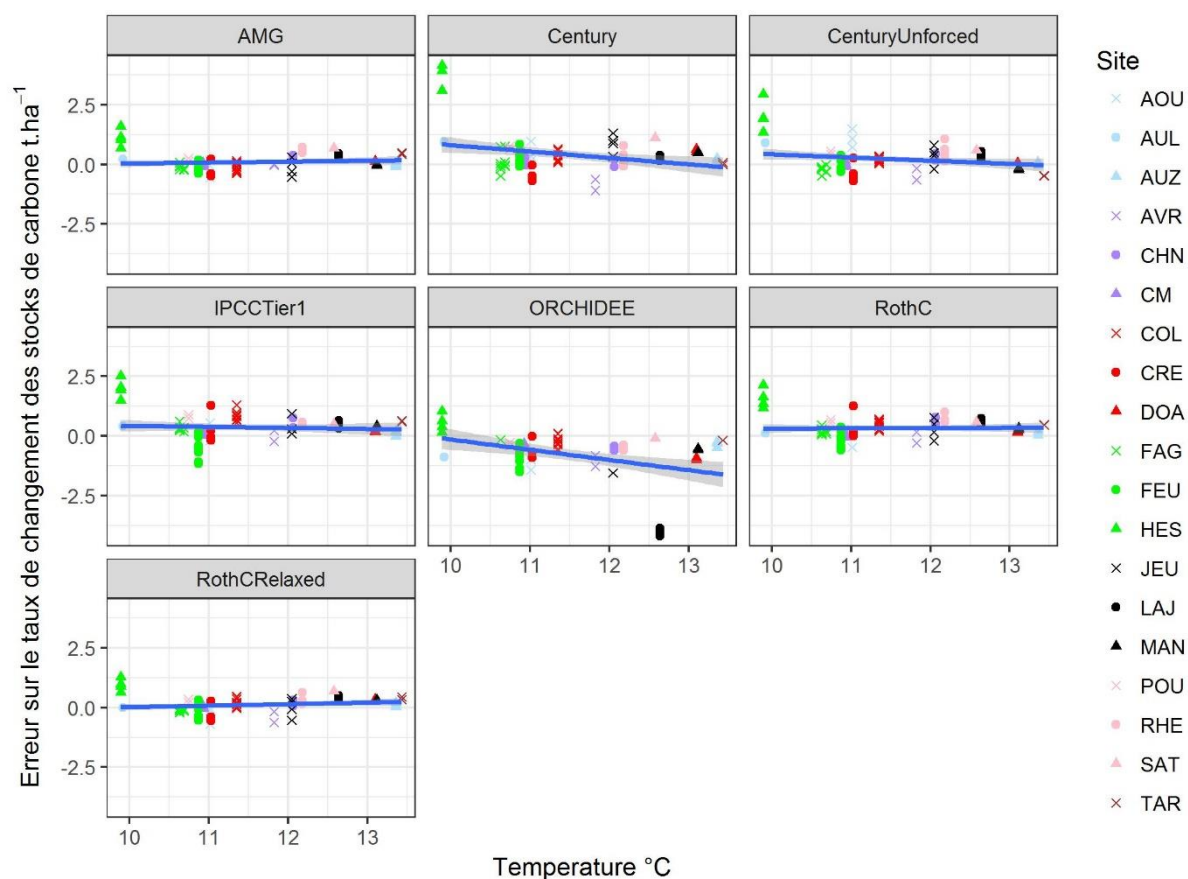


Figure 8 :Erreur sur le taux de changement des stocks de carbone des modèles en fonction de la température.

Tableau 8: Test sur la pente et l'ordonnée à l'origine des modèles en fonction de la température. Les cases rouges pour les p-valeur < 5% (test significatif)

Modèle	Ordonnée à l'origine estimée	p-valeur	Pente de la droite estimée	p-valeur	R ²
AMG	0.0954414	0.86	0.0002219	0.996	-0.01
Century	4.6904	0.000229	-0.3697	0.00072	0.136
Century Unforced	2.55459	0.00189	-0.1997	0.00475	0.0931
ORCHIDEE	4.1236	0.004980	-0.4273	0.000886	0.1313
IPCC Tier1	1.11293	0.233	-0.06634	0.41	-0.004316
RothC	0.58437	0.395	-0.0228	0.7	-0.01179
RothC Relaxed	-0.26642	0.608	0.03343	0.457	-0.006073

La variable « Température » prend des valeurs comprises entre 9.89°C et 11.45°C, un intervalle pas aussi large pour que la variation soit significative mais on remarque que les modèles Century, Century Unforced et ORCHIDEE sont sensibles à la température.

Bilan hydrique (P-ETP) :

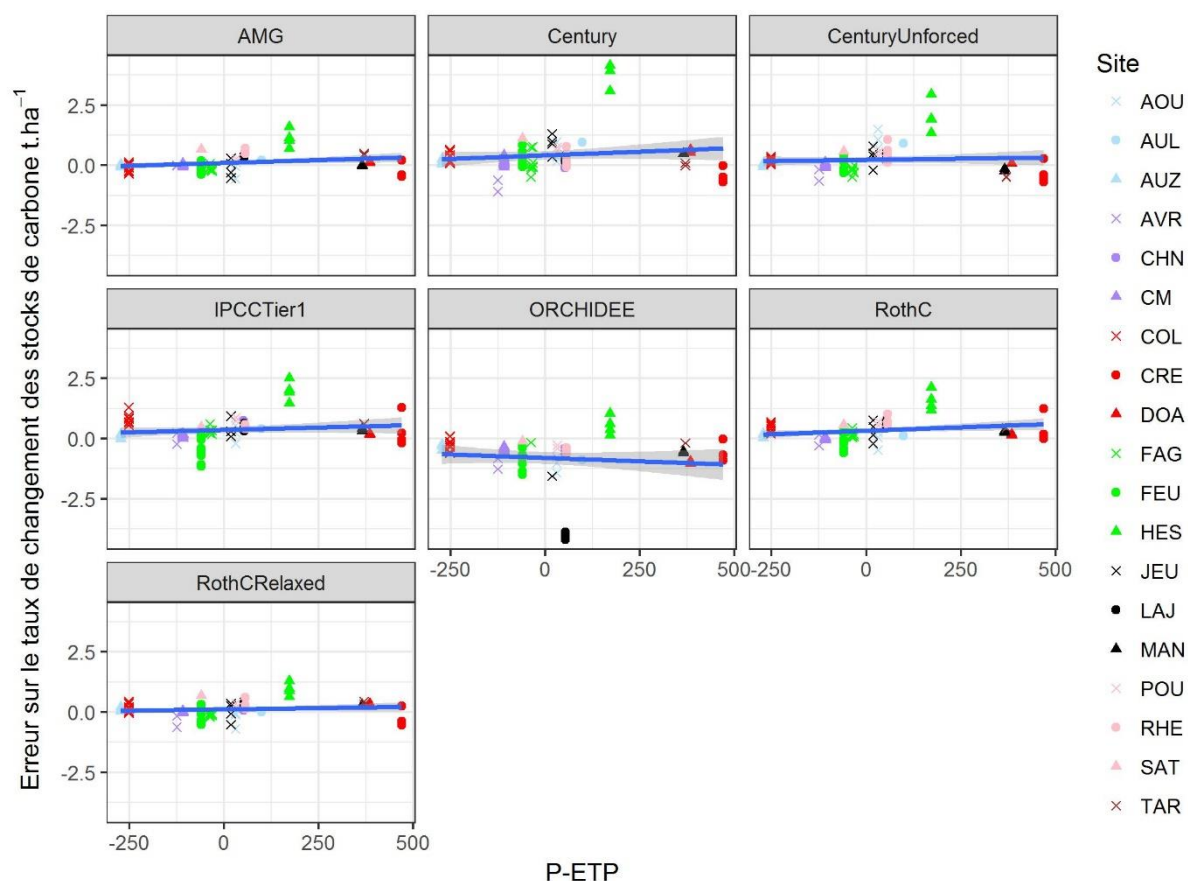


Figure 9: :Erreur sur le taux de changement des stocks de carbone des modèles en fonction du bilan hydrique. P-ETP : Précipitation- évapotranspiration potentielle (mm). La droite de régression et son intervalle de confiance à 95%

Tableau 9: Test sur la pente et l'ordonnée à l'origine des modèles en fonction du bilan hydrique. Les cases rouges pour les p-valeur < 5% (test significatif)

Modèle	Ordonnée à l'origine estimée	p-valeur	Pente de la droite de régression estimée	p-valeur	R ²
AMG	0.1027333	0.0183	0.0004173	0.0517	0.03838
Century	0.4396019	0.0000793	0.0006275	0.232	0.006167
Century Unforced	0.2575004	0.000289	0.0002380	0.479583	-0.006826
ORCHIDEE	-0.8033189	9.52x10 ⁻⁹	-0.0005713	0.355	-0.001826
IPCC Tier1	0.3532	1.11x10 ⁻⁵	0.0003790	0.31	0.0006347
RothC	0.3280303	4.69x10 ⁻⁸	0.0005537	0.0412	0.04354
RothC Relaxed	0.1202615	0.00531	0.0001522	0.46537	-0.00636

Pour tous les modèles, les ordonnées à l'origine sont significativement différentes de zéro. Les modèles AMG et RothC Relaxed restent quand même moins sensibles au bilan hydrique.

Argile :

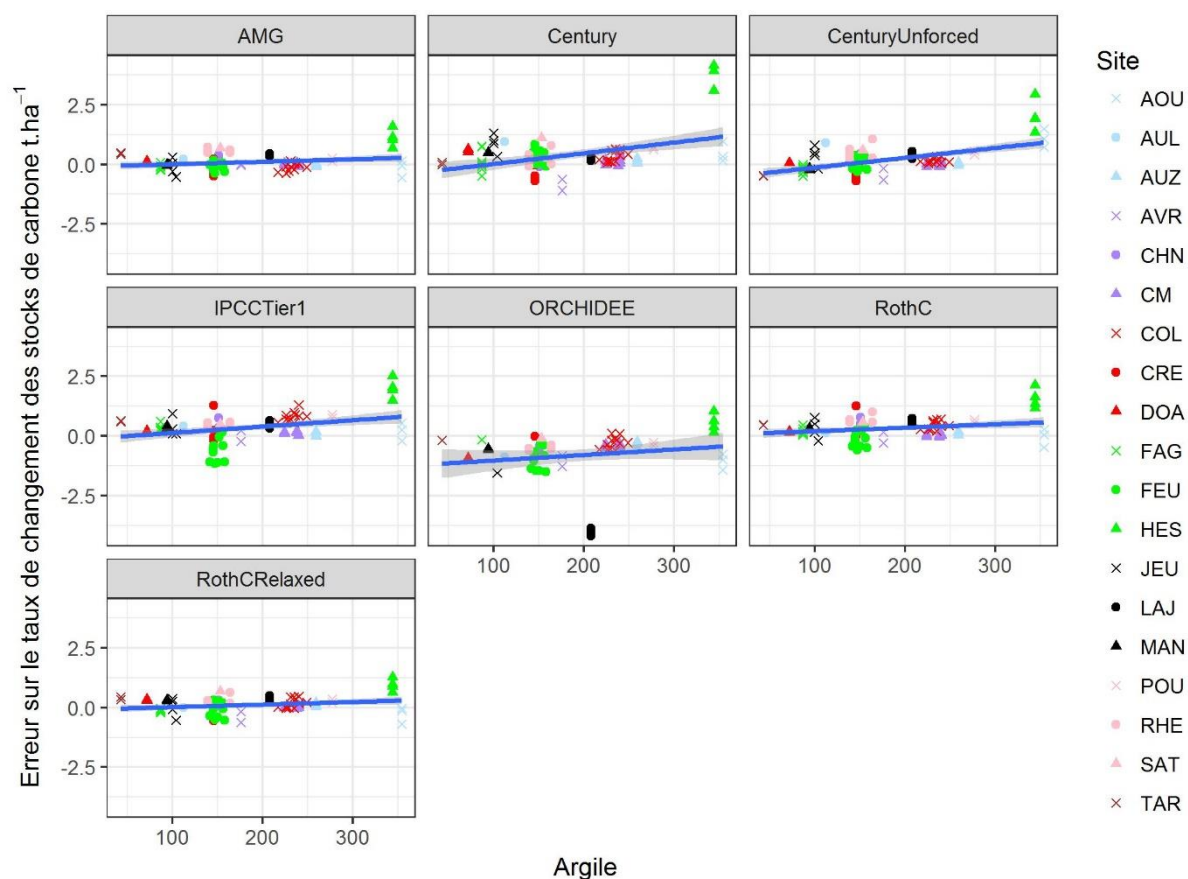


Figure 10 :Erreur sur le taux de changement des stocks de carbone des modèles en fonction de l'argile (g. Kg-1). La droite de régression et son intervalle de confiance à 95%

Tableau 10: Test sur la pente et l'ordonnée à l'origine des modèles en fonction de l'argile. Les cases rouges pour les p-valeur < 5% (test significatif)

Modèle	Ordonnée à l'origine estimée	p-valeur	Pente de la droite de régression estimée	p-valeur	R ²
AMG	-0.1263612	0.3274	0.00111105	0.0675	0.03242
Century	-0.71	0.14	0.005686	5.67x10 ⁻⁵	0.19
Century Unforced	-0.6992985	6.59x10 ⁻⁵	0.0047	4.22x10 ⁻⁶	0.3339
ORCHIDEE	-1.2632	0.00103	0.0023	0.18476	0.01075
IPCC Tier1	-0.4334864	0.036866	0.0038728	0.000119	0.1758
RothC	-0.0731032	0.6462	0.0019543	0.0102	0.07561
RothC Relaxed	-0.1250710	0.3100	0.0012057	0.0382	0.04523

Seul le modèle AMG est de pente nul et d'erreur très faibles à l'origine. L'erreur sur le taux de changement des stocks de carbone pour les autres modèles varie avec la teneur en argile. Surtout Century Unforced.

Calcaire :

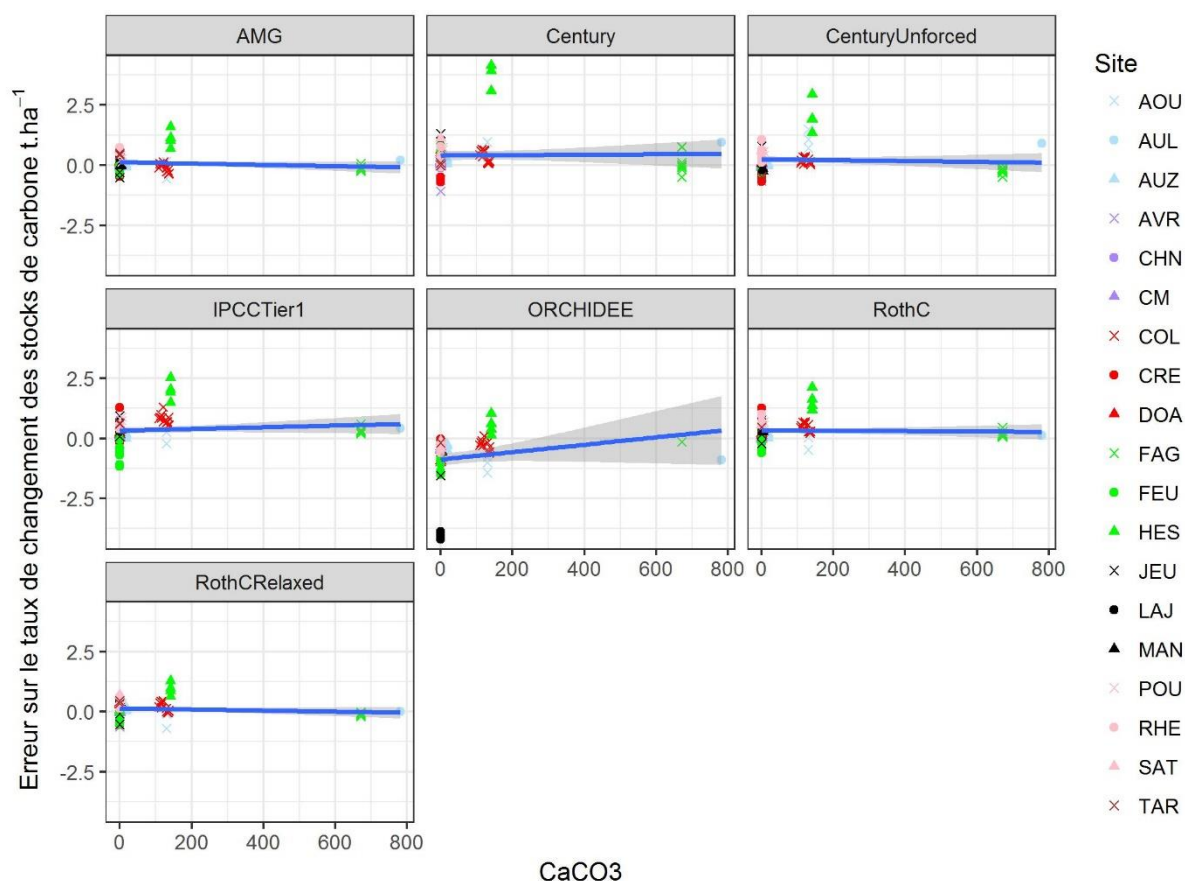


Figure 11: Erreur sur le taux de changement des stocks de carbone des modèles en fonction de CaCO_3 (g. Kg^{-1}). La droite de régression et son intervalle de confiance à 95%

Tableau 11: Test sur la pente et l'ordonnée à l'origine des modèles en fonction du Calcaire. Les cases rouges pour les p-valeur < 5% (test significatif)

Modèle	Ordonnée à l'origine estimée	p-valeur	Pente de la droite de régression estimée	p-valeur	R ²
AMG	0.0929	0.0541	9.47×10^{-5}	0.7859	-0.01284
Century	0.3227	0.00465	0.00203	0.01398	0.068
Century Unforced	0.1878	0.00989	0.0012435	0.0192	0.0608
ORCHIDEE	-0.8798	5.6×10^{-10}	0.001543	0.118	0.02025
IPCC Tier1	0.2804	0.000702	0.00127	0.0315	0.0496
RothC	0.30164	3.28×10^{-10}	0.0003735	0.397	-0.003755
RothC Relaxed	0.1088759	0.0194	0.0001795	0.5922	-0.009828

A part les deux traitements dans les sites Fagnières (FAG_L9) et Aulnay sous-bois (AUL_11) qui sont des sols très calcaires, les autres traitements sont effectués dans des sols à très proches valeurs de CaCO_3 . Tous les modèles à part AMG sont sensibles à cette variable.

pH :

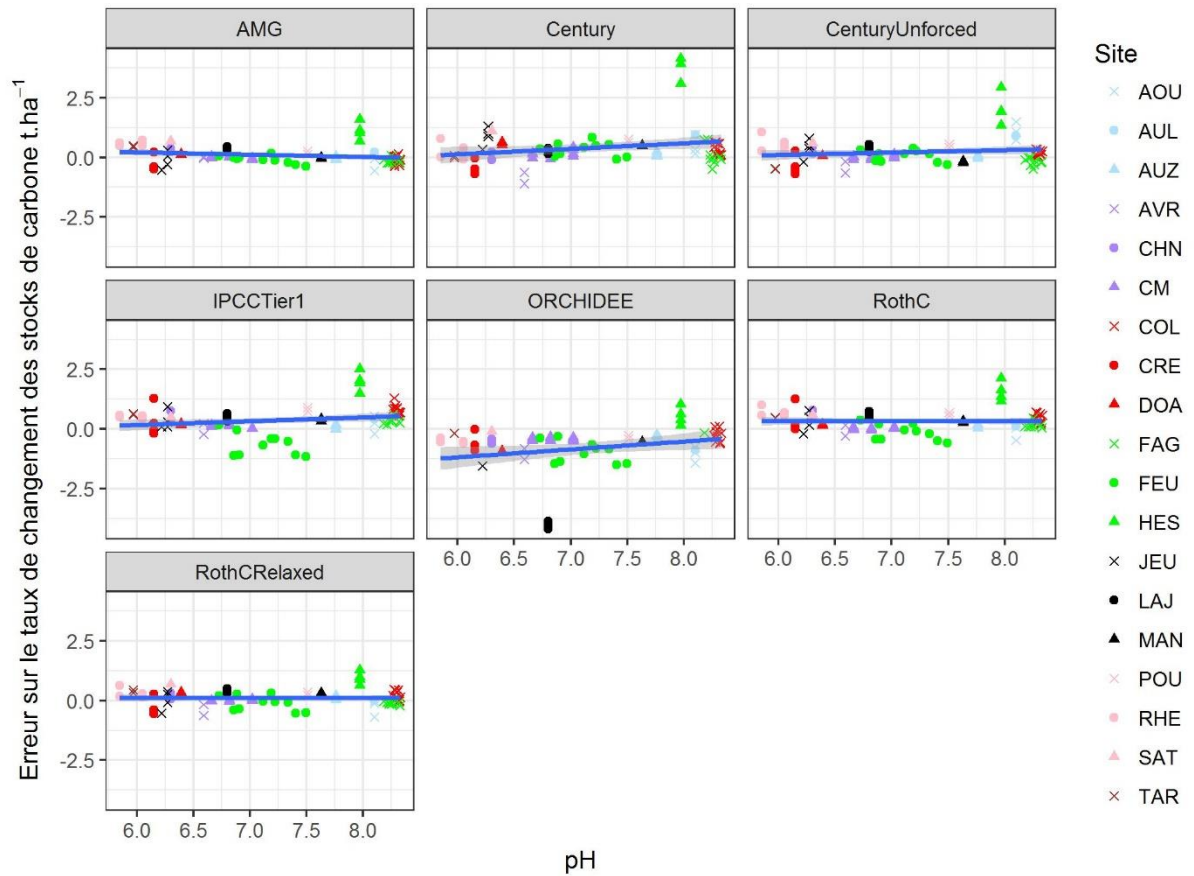


Figure 12: :Erreur sur le taux de changement des stocks de carbone des modèles en fonction du pH. La droite de régression et son intervalle de confiance à 95%

Tableau 12: Test sur la pente et l'ordonnée à l'origine des modèles en fonction du pH. Les cases rouges pour les p-valeur < 5% (test significatif)

Modèle	Ordonnée à l'origine estimée	p-valeur	Pente de la droite de régression estimée	p-valeur	R ²
AMG	0.52424	0.195	-0.05929	0.288	0.002008
Century	-2.5166	0.00753	0.4102	0.0018	0.1152
Century Unforced	-1.3584	0.02569	0.2244	0.00815	0.08068
ORCHIDEE	-3.1636	0.00592	0.3292	0.03620	0.4646
IPCC Tier1	-1.2635	0.0628	0.2243	0.0178	0.06273
RothC	0.001871	0.997	0.044495	0.53	-0.008302
RothC Relaxed	-0.33262	0.389	0.06276	0.241	0.005396

Nous remarquons que l'erreur des modèles reste très faible quelle que soit la valeur du pH pour certains modèles. Cependant le test sur la régression montre que l'erreur varie en fonction du pH pour tous les modèles sauf AMG.

Stock initial de carbone :

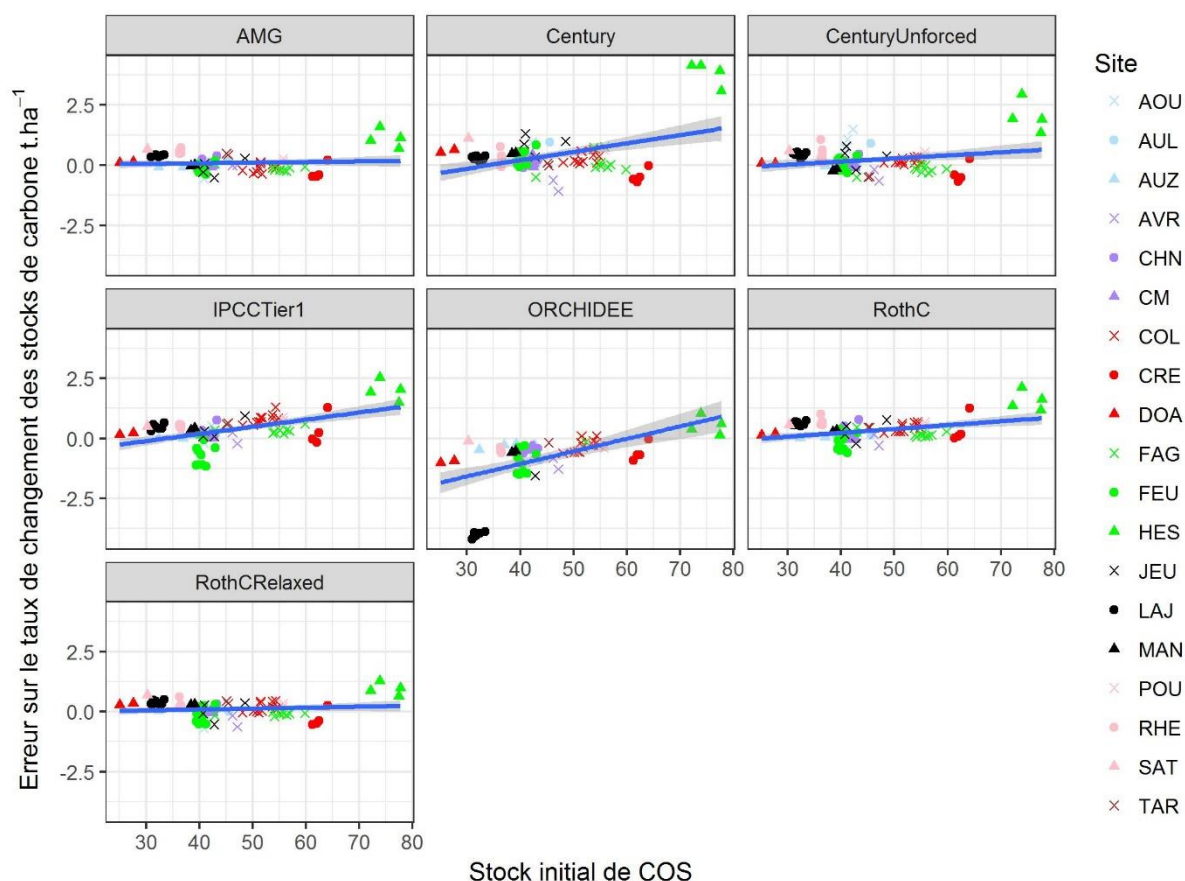


Figure 13: :Erreur sur le taux de changement des stocks de carbone des modèles en fonction du stock initial de carbone. La droite de régression et son intervalle de confiance à 95%

Tableau 13: Test sur la pente et l'ordonnée à l'origine des modèles en fonction du stock initial de carbone. Les cases rouges pour les p-valeur < 5% (test significatif)

Modèle	Ordonnée à l'origine estimée	p-valeur	Pente de la droite de régression estimée	p-valeur	R ²
AMG	-0.1668	0.361	0.005887	0.137	0.01694
Century	-1.47	0.000256	0.042	2.43x10 ⁻⁶	0.25
Century Unforced	-0.59	0.03	0.018	0.0017	0.1163
ORCHIDEE	-3.139	6.94x10 ⁻¹⁰	0.052	6.43x10 ⁻⁷	0.283
IPCC Tier1	-1.149	3.48x10 ⁻⁵	0.0333	9.88x10 ⁻⁶	0.3184
RothC	-0.58	0.0056	0.0201	2.14x10 ⁻⁵	0.2124
RothC Relaxed	-0.17	0.3232	0.0064	0.0885	0.02644

Le test n'accepte que les modèles AMG et RothC Relaxed, les autres modèles s'avèrent imprécis en fonction du stock initial de carbone. Comme les pentes sont positives, ces modèles auront tendance à surestimer le taux de stockage de carbone pour des traitements à stock initial de carbone important.

Fertilisants organiques :

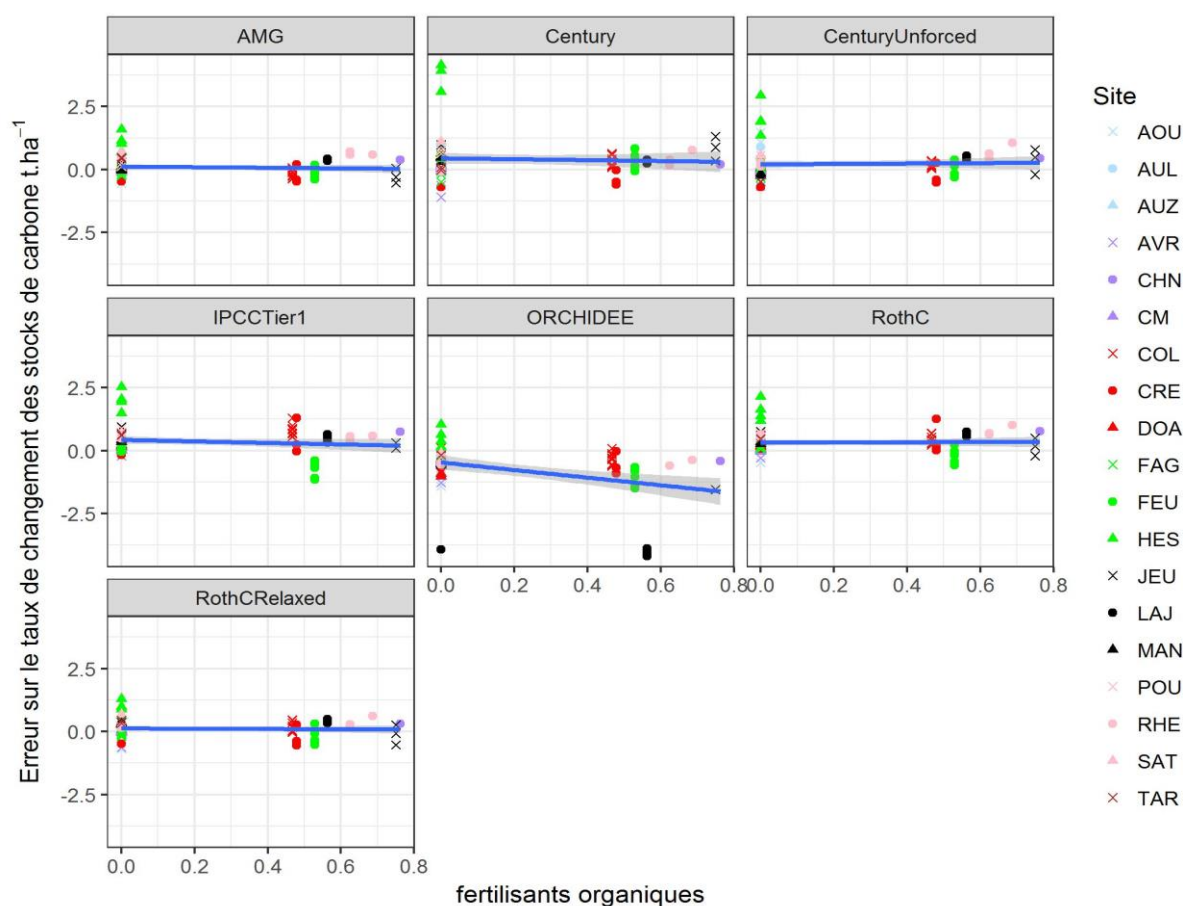


Figure 14: Erreur sur le taux de changement des stocks de carbone des modèles en fonction de l'apport de fertilisants organiques. La droite de régression et son intervalle de confiance à 95%

Tableau 14: Test sur la pente et l'ordonnée à l'origine des modèles en fonction de l'apport des fertilisants organiques. Les cases rouges pour les p-valeur < 5% (test significatif)

Modèle	Ordonnée à l'origine estimée	p-valeur	Pente de la droite de régression estimée	p-valeur	R ²
AMG	0.14	0.010	-0.22	0.168	0.012
Century	0.53	0.000156	-0.49	0.216	0.0075
Century Unforced	0.29	0.00101	-0.19	0.43	-0.0053
ORCHIDEE	-0.47	0.0021	-1.50	0.000833	0.13
IPCC Tier1	0.41	4.16x10 ⁻⁵	-0.32	0.252	0.0045
RothC	0.33	1.33x10 ⁻⁵	0.049	0.813	-0.013
RothC Relaxed	0.15	0.0065	-0.14	0.3531	-0.001

Nous remarquons que beaucoup de traitements sont à zéro ajout de fertilisants organiques. Le test rejette l'hypothèse nulle pour tous les modèles. L'erreur est donc très sensible à cette pratique agricole.

Travail du sol :

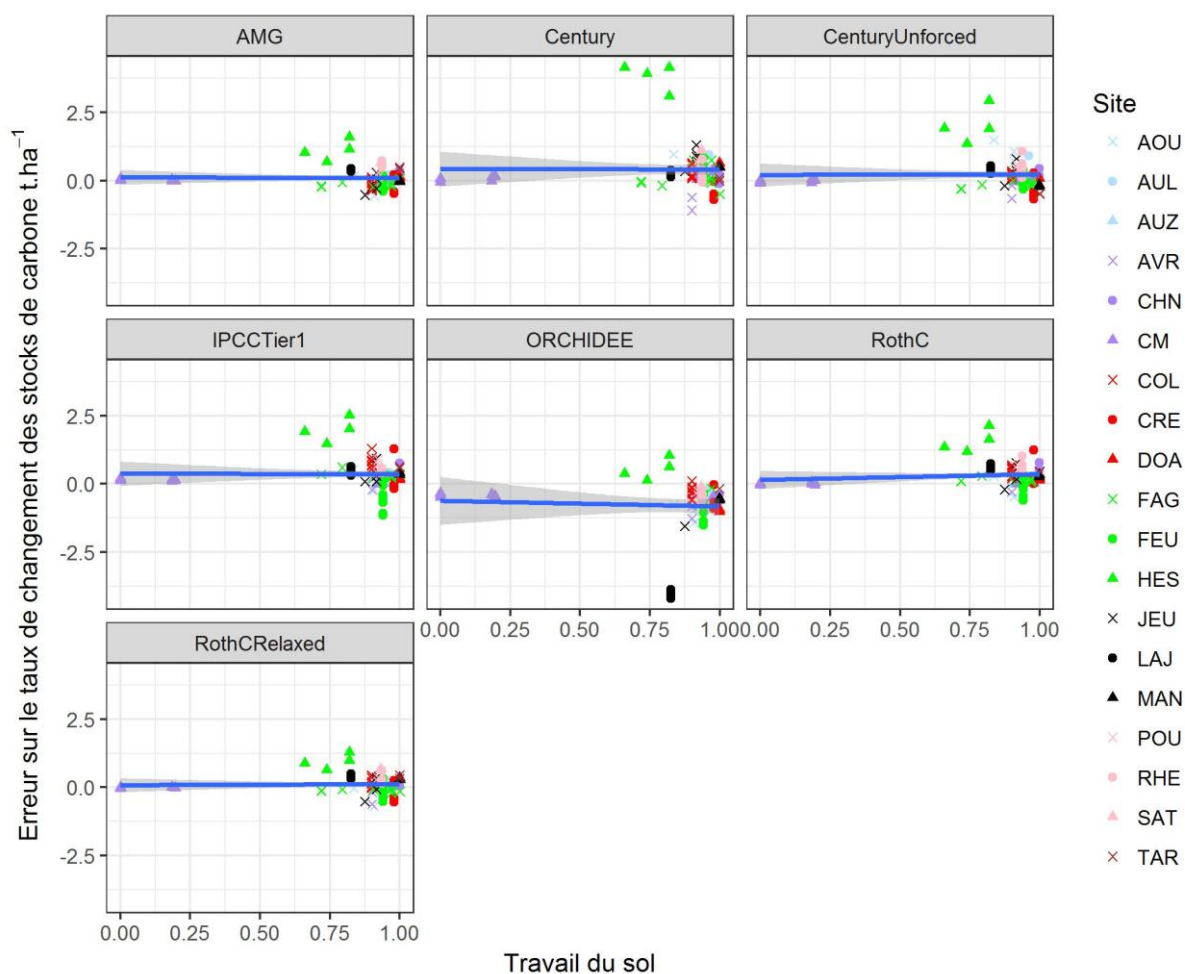


Figure 15: Erreur sur le taux de changement des stocks de carbone des modèles en fonction du travail du sol. La droite de régression et son intervalle de confiance à 95%

Tableau 15: Test sur la pente et l'ordonnée à l'origine des modèles en fonction du travail des sols. Les cases rouges pour les p-valeur < 5% (test significatif)

Modèle	Ordonnée à l'origine estimée	p-valeur	Pente de la droite de régression estimée	p-valeur	R ²
AMG	0.14	0.36	-0.05	0.76	-0.012
Century	0.49	0.19	-0.078	0.85	-0.013
Century Unforced	0.24	0.31	0.012	0.96	-0.013
ORCHIDEE	-0.62	0.163	-0.2065	0.68	-0.011
IPCC Tier1	0.39	0.14	-0.058	0.848	-0.013
RothC	0.17	0.37	0.17	0.43	-0.0054
RothC Relaxed	0.095	0.526	0.027	0.87	-0.013

Le test n'est pas significatif pour tous les modèles. On peut conclure que l'erreur ne dépend pas du travail du sol.

Les résidus des cultures :

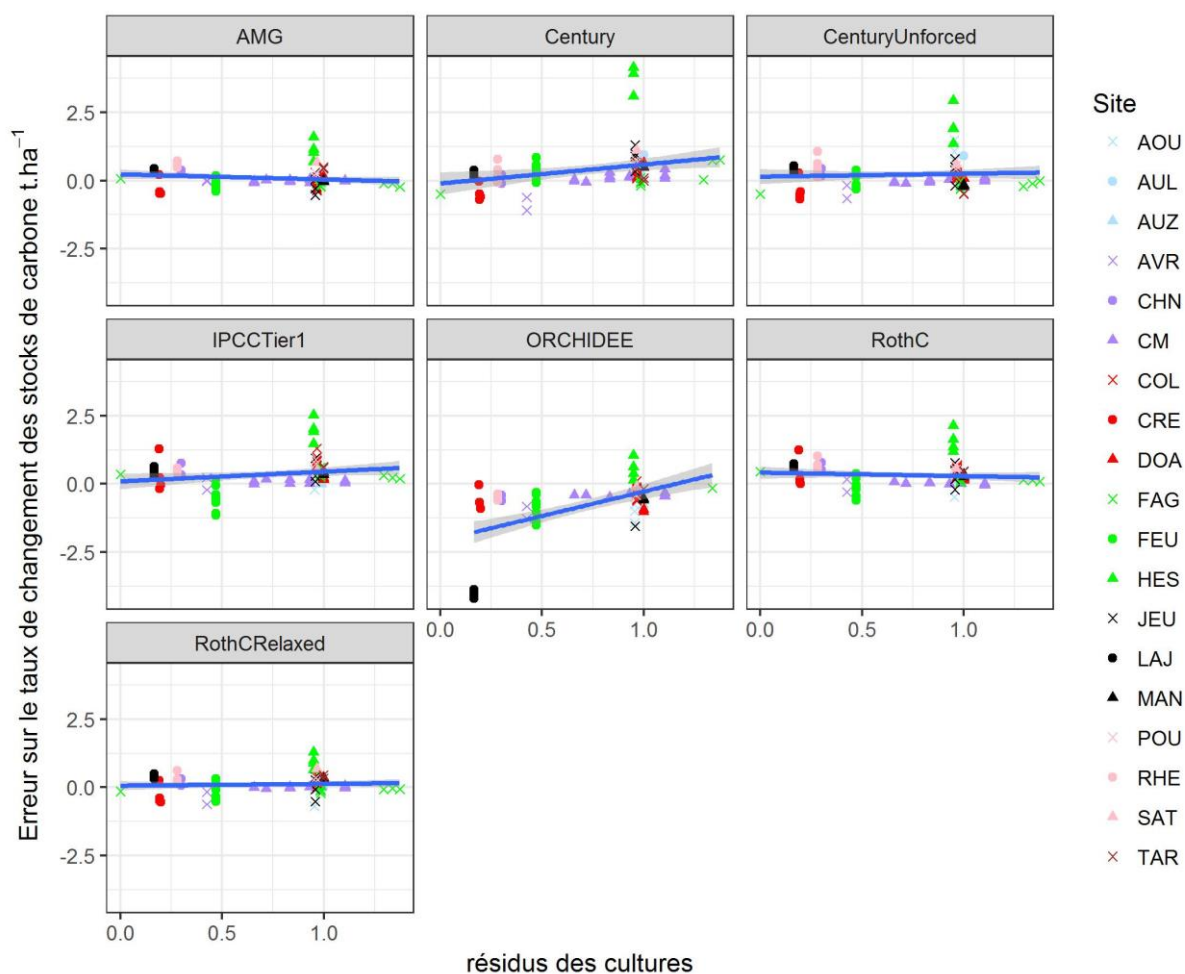


Figure 16: :Erreur sur le taux de changement des stocks de carbone des modèles en fonction des résidus des cultures La droite de régression et son intervalle de confiance à 95%

Tableau 16: Test sur la pente et l'ordonnée à l'origine des modèles en fonction des résidus animaux et végétaux. Les cases rouges pour les p-valeur < 5% (test significatif)

Modèle	Ordonnée à l'origine estimée	p-valeur	Pente de la droite de régression estimée	p-valeur	R ²
AMG	0.17	0.0845	-0.11	0.3839	-0.0031
Century	-0.17	0.46	0.85	0.00523	0.086
Century Unforced	0.088	0.579	0.23	0.25	0.0047
ORCHIDEE	-2.06	1.82x10 ⁻¹²	1.79	1.93x10 ⁻⁷	0.30
IPCC Tier1	0.014	0.93	0.47	0.0355	0.046
RothC	0.36	0.00607	-0.066	0.69	-0.011
RothC Relaxed	0.019	0.84	0.14	0.26	0.0033

Les résidus des cultures ont tendance à augmenter le stock de carbone. Les modèles AMG, Century Unforced et RothC Relaxed sont très précis. Seul le modèle ORCHIDEE est sensible aux résidus.

Cultures intermédiaires :

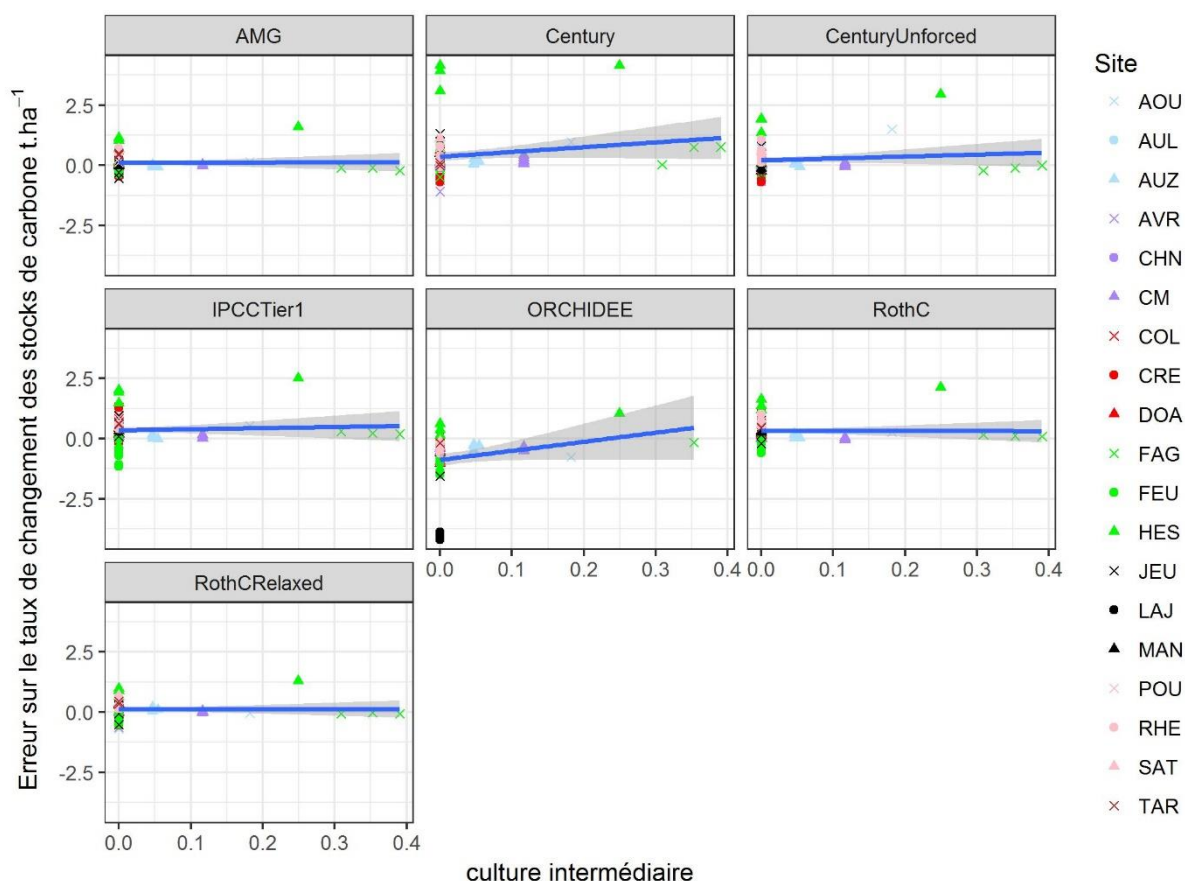


Figure 17: Erreur des modèles en fonction des cultures intermédiaires. La droite de régression et son intervalle de confiance à 95%

Tableau 17: Test sur la pente et l'ordonnée à l'origine des modèles en fonction des cultures intermédiaires. Les cases rouges pour les p-valeur < 5% (test significatif)

Modèle	Ordonnée à l'origine estimée	p-valeur	Pente de la droite de régression estimée	p-valeur	R ²
AMG	0.08	0.086	0.778	0.27	0.0025
Century	0.35	0.0018	3.35	0.052	0.038
Century Unforced	0.208	0.0042	2.06	0.062	0.034
ORCHIDEE	-0.88	2.47 x10 ⁻¹¹	3.77	0.063	0.033
IPCC Tier1	0.323	0.00012	1.14	0.35	-0.0019
RothC	0.31	1.21x10 ⁻⁶	0.43	0.630	-0.0106
RothC Relaxed	0.107	0.018	0.47	0.49	-0.0073

Toutes les pentes de régression sont quasiment nulles sauf pour le modèle ORCHIDEE. Donc les modèles prennent bien en compte cette pratique et l'erreur n'en dépend pas.

Culture de Colza :

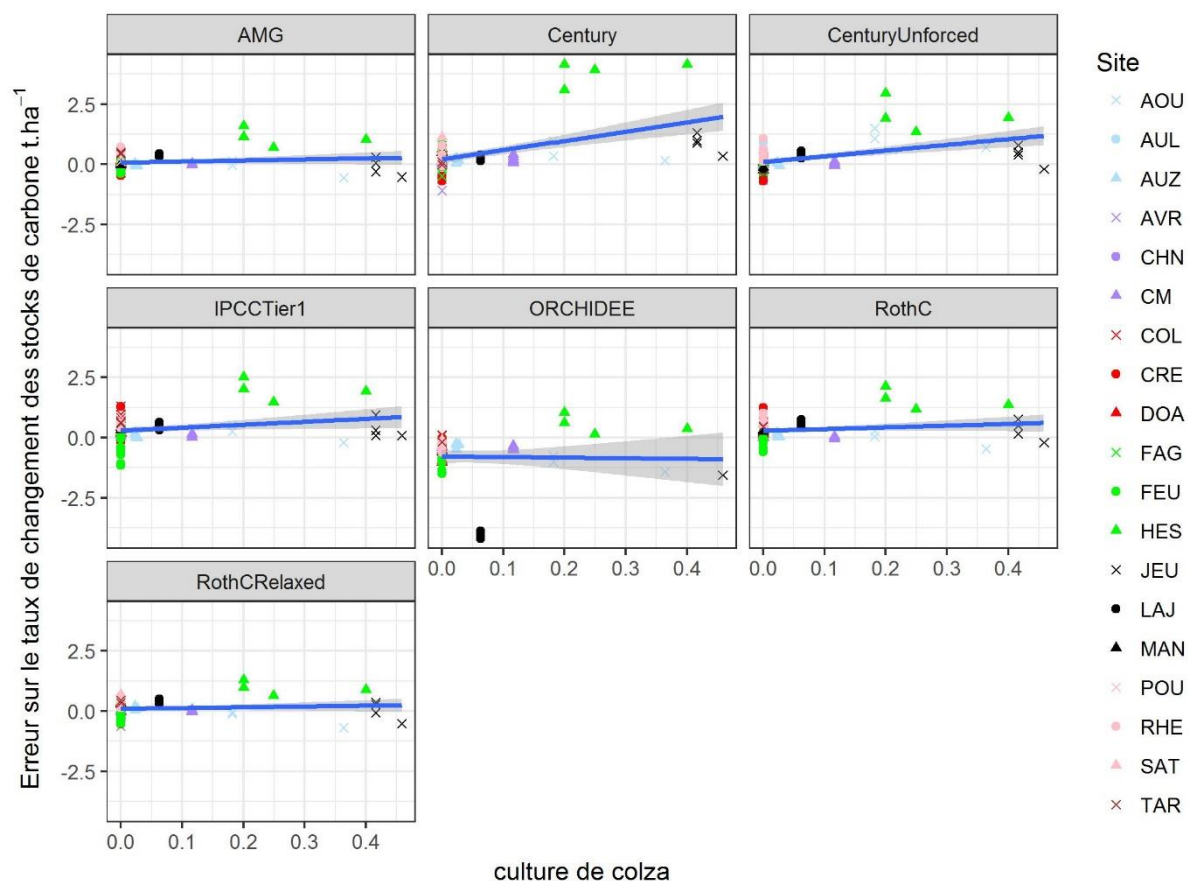


Figure 18::Erreur des modèles en fonction des cultures de colza. La droite de régression et son intervalle de confiance à 95%

Tableau 18: Test sur la pente et l'ordonnée à l'origine des modèles en fonction des culture de Colza. Les cases rouges pour les p-valeur < 5% (test significatif)

Modèle	Ordonnée à l'origine estimée	p-valeur	Pente de la droite de régression estimée	p-valeur	R ² ajusté
AMG	0.062	0.1919	0.768	0.094	0.025
Century	0.20	0.048	5.03	1.25x10 ⁻⁶	0.27
Century Unforced	0.113	0.085	3.07	4.28x10 ⁻⁶	0.24
ORCHIDEE	-0.78	2.66x10 ⁻⁷	-0.25	0.85	-0.013
IPCC Tier1	0.26	0.0016	1.86	0.016	0.063
RothC	0.28	1.29x10 ⁻⁵	0.82	0.156	0.014
RothC Relaxed	0.104	0.0278	0.309	0.4868	-0.007

Le modèle Century et Century Unforced sont très sensibles à la culture de Colza, RothC et RothC Relaxed un peu moins. Le modèle AMG reste très précis et prend bien en compte cette pratique agricole.

Synthèse: Le tableau récapitulatif des modèles et les variables pour lesquelles le test est significatif

Tableau 19: Tableau récapitulatif des tests sur les pentes des droites de régression. (-) test non significatif. (x) test significatif.

Modèle de simulation	T (°C)	P-ETP (mm/an)	Argile g/Kg	CaCO ₃ g/Kg	pH	Soc initial t/ha	F.O	TdS	R.C	C.I	Colza
AMG	-	-	-	-	-	-	-	-	-	-	-
Century	x	-	x	x	x	x	-	-	x	-	x
Century Unforced	x	-	x	x	x	x	-	-	-	-	x
ORCHIDEE	x	x	-	-	x	x	x	-	x	-	-
IPCC Tier1	-	-	x	x	x	x	-	-	x	-	x
RothC	-	-	x	-	-	x	-	-	-	-	-
RothC Relaxed	-	-	x	-	-	-	-	-	-	-	-

T: Température;

P-ETP: Bilan hydrique;

CaCO₃: Calcaire;

Soc: Stock de carbone organique;

F.O: fertilisants organiques;

TdS: travail du sol;

R.C: Résidus des cultures;

C.I: cultures intermédiaires;

Colza: Culture de Colza

Comme on s'intéresse à la sensibilité de l'erreur par rapport aux variables pédoclimatiques et aux pratiques agricoles, on se focalise sur la pente de régression. Nous remarquons d'après le tableau ci-dessus (tableau 19), que le modèle AMG est très précis et n'est sensible à aucune variable. L'erreur du modèle RothC Relaxed est en corrélation uniquement avec les sols argileux. L'erreur de RothC est sensible à l'argile et au stock initial de carbone. Le modèle Century, même s'il est plus précis que la plupart des modèles, il réagit rapidement à la variation de température, de teneur en calcaire et en argile, du pH du stock initial de carbone des résidus et de la culture de Colza.

Seul le modèle ORCHIDEE est sensible au bilan hydrique et à l'ajout de fertilisants organiques alors que toutes les erreurs des modèles tiennent en compte le travail du sol (labour).

4.3. Analyse multivariée

Dans cette partie nous allons nous intéresser à l'étude d'éventuelles corrélations entre l'erreur des modèles et les variables à l'aide d'une approche multivariée.

L'histogramme d'éboulis des valeurs propres qui retourne le pourcentage d'inertie retenu pour chaque axe de l'ACP :

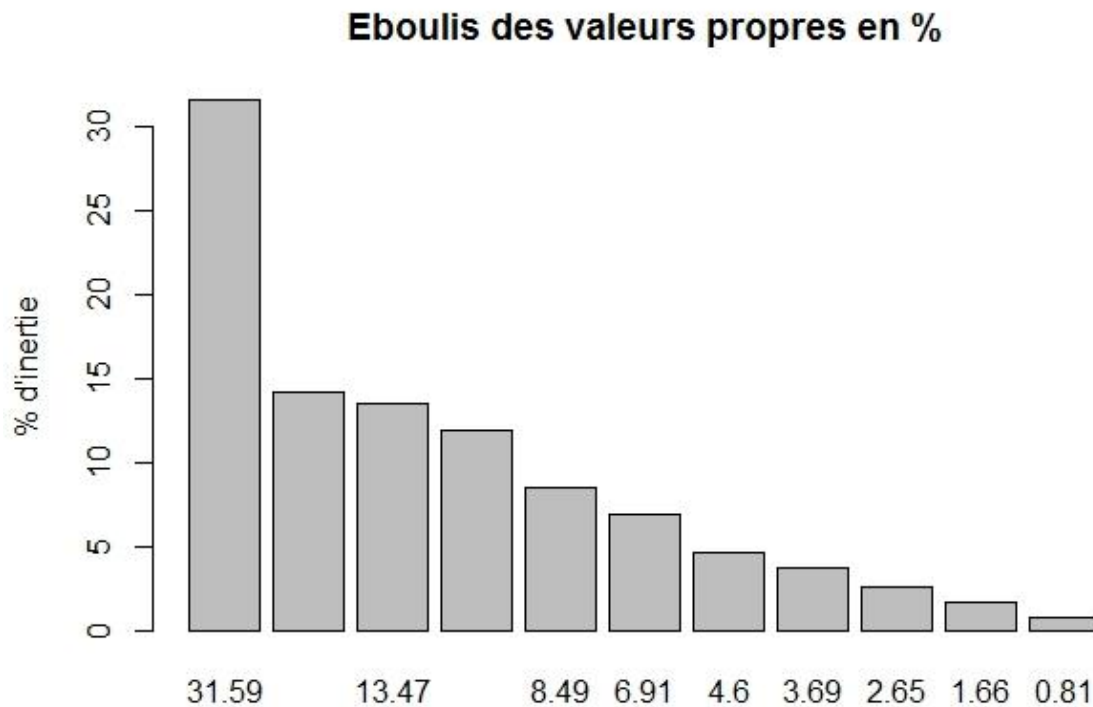


Figure 19: histogramme d'éboulis des valeurs propres

Deux axes retenus pour l'analyse. A l'aide du Package (factoextra), la double projection (cercle de corrélation et le plan des individus) sur la même figure ci-dessous (Figure 20) dans le cas où les erreurs des modèles ne sont pas incluses dans le calcul de l'ACP.

Tableau 19: Tableau des \cos^2 pour les deux premières dimensions.

	Dim.1	Dim.2
Température	0.36658878	0.07655896
P.ETP	0.14137680	0.29933361
Argile	0.51650462	0.02274674
CaCO ₃	0.32780158	0.11458888
pH	0.57645123	0.29363936
Fertilisants organiques	0.11993212	0.09830659
Travail du sol	0.07921333	0.26472386
Résidus des cultures	0.52968066	0.05026903
Cultures intermédiaires	0.24847988	0.02568625
Culture de colza	0.21003871	0.27542101
SOC initial	0.35903724	0.04310559

Dans un deuxième temps nous incluons l'erreur dans les calculs de l'ACP :

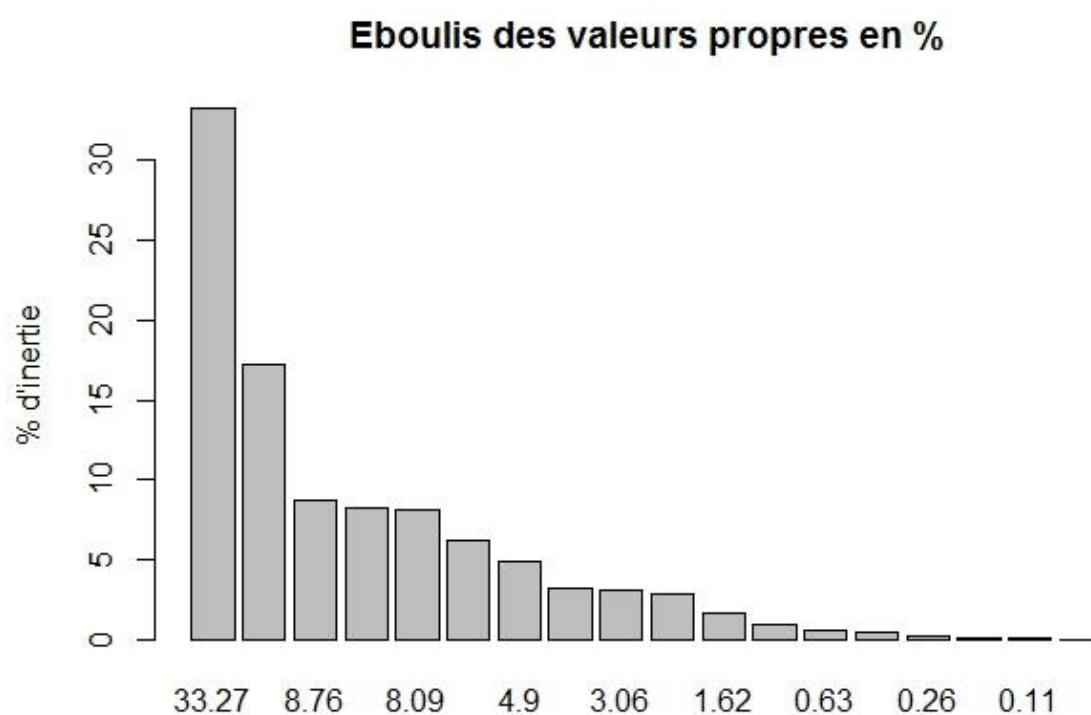


Figure 21: histogramme d'éboulis des valeurs propres

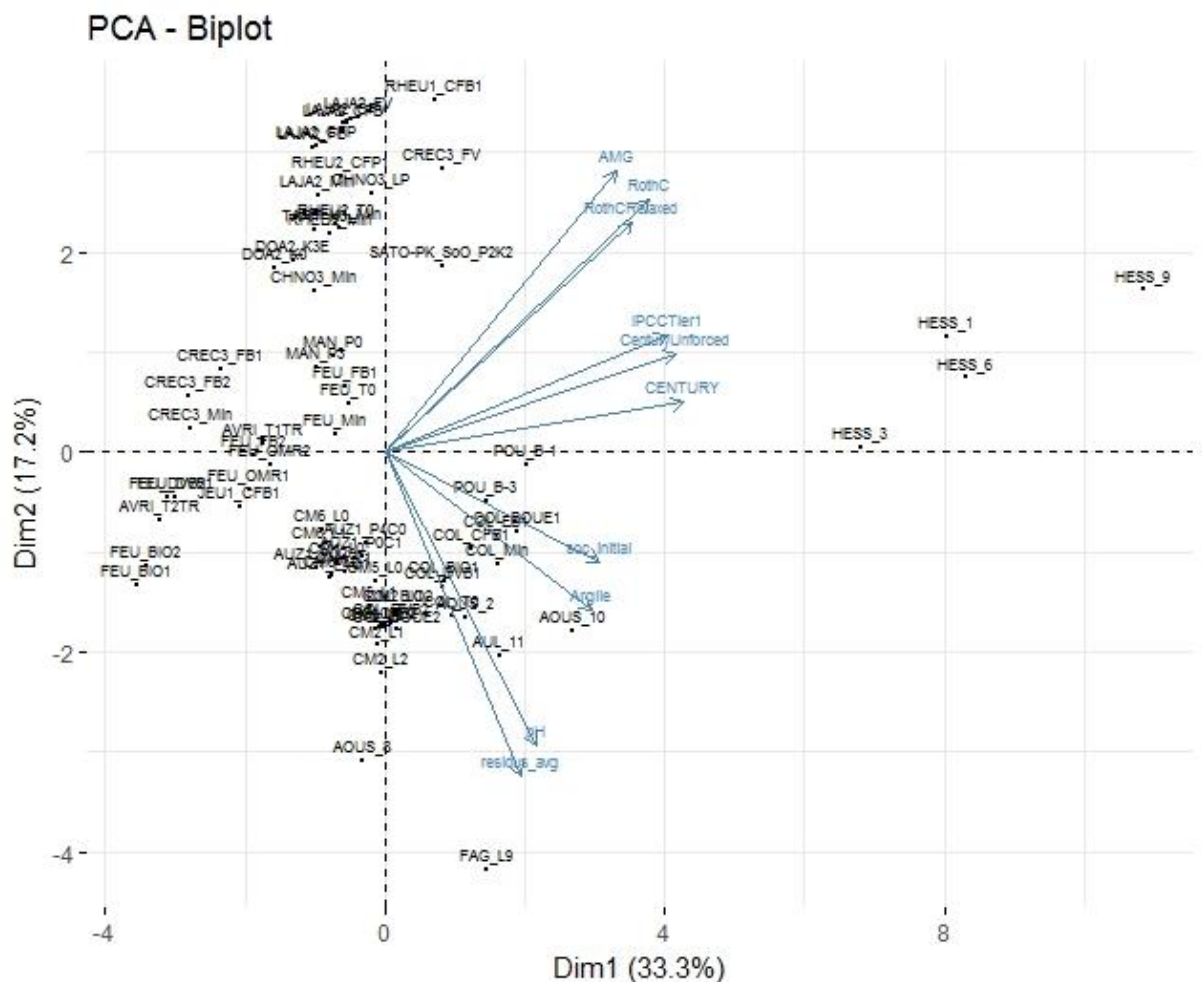


Figure 22: Cercle de corrélation de l'ACP avec les erreurs des modèles incluses et plan des individus (Traitements)

Dans ce cas, où les erreurs des modèles sont incluses dans les calculs de l'ACP, nous remarquons une forte corrélation entre l'erreur des modèles AMG, RothC et RothC Relaxed qui sont les plus précis et qui contribuent à la formation aussi bien de l'axe 1 que l'axe 2. Aussi une forte corrélation entre Century, Century Unforced et IPCC Tier1. Le modèle ORCHIDEE, quant à lui est mal représenté et n'est pas projeté sur le plan. En revanche le pourcentage d'inertie retenu reste très faible (33.3% pour le premier axe et 17.2% pour le deuxième). Nous remarquons aussi que le site HESS (Hessange) avec ces différents traitements représente des erreurs importantes suivant tous les modèles. Ci-dessous le tableau des \cos^2 des variables et des traitements pour les 12 dimensions. Plus la valeur de \cos^2 est proche de 1, meilleure est la qualité de représentation. Donc l'axe 1 est un axe formé par les erreurs des modèles sauf ORCHIDEE.

Tableau 20: Tableau des \cos^2 pour les deux premières dimensions.

	Dim.1	Dim.2
AMG	0.480936715	0.34750721
CENTURY	0.801465825	0.01096224
Century Unforced	0.757631440	0.04268577
IPCCTier1	0.713710811	0.06063763

ORCHIDEE	0.173771147	0.17363197
RothC	0.626525757	0.27844528
RothC Relaxed	0.550228419	0.23172768
Température	0.163345901	0.20379186
P-ETP	0.000506767	0.30467346
Argile	0.385382158	0.10897694
CaCO3	0.143659988	0.17892361
pH	0.205785523	0.37812779
Fertilisants organiques	0.067694114	0.06213673
Travail du sol	0.009181962	0.06256519
Résidus des cultures	0.167981834	0.45856029
Cultures intermédiaires	0.103812818	0.12633838
Culture de Colza	0.221364817	0.01075069
SOC initial	0.415039149	0.05245947

5. Conclusion :

L'objectif de ce stage était de déterminer les facteurs qui influent l'erreur du taux de changement des stocks de carbone et de classer les modèles de simulation selon leur précision. Afin de répondre aux objectifs, j'ai utilisé deux méthodes: le test de Kruskal-Wallis et la régression linéaire simple de l'erreur en fonction de chaque variable, qui a permis de voir les facteurs influant cette erreur.

La première méthode a montré que le modèle AMG est très précis suivi de RothC et Century au deuxième rang. Ensuite les modèles RothC Relaxed et Century Unforced viennent au troisième rang suivi du modèle statistique IPCC Tier1 et enfin le modèle ORCHIDEE qui montre les plus grandes valeurs de l'erreur. La régression linéaire simple de l'erreur a montré que l'erreur sur le taux de changement des stocks de carbone est plus sensible aux variables liées aux sites comme l'argile et le stock initial de carbone, et moins sensible aux pratiques agricoles comme le travail du sol. A noter une grande incertitude dans les mesures comme c'est le cas dans les sites Hessange et Feucherolles.

En conclusion, ce stage m'a permis d'apprendre les bases de l'étude statistique et de mettre en application le test de régression linéaire ainsi que l'analyse en composante principale. Il m'a également montré l'importance du prétraitement des données et l'étude des distributions de toutes les variables.

J'ai pu approfondir mes connaissances en programmation R et découvrir de nouveaux outils (Packages), principalement avec l'interface R-Studio et le Package (ggplot2 et FactoMineR). J'ai également appris à travailler en équipe et mener des études statistiques.

Plus généralement, les codes R que j'ai mis en place pourront permettre de tester l'effet de certaines pratiques sur les modèles de simulation sur des bases de données plus larges.

6. Liste des figures:

- FIGURE 1: ORGANIGRAMME DE L'UNITE INFOSol
- FIGURE 2: CARTE GEOGRAPHIQUE DES SITES D'ESSAIS
- FIGURE 3: HISTOGRAMMES: DISTRIBUTION DES VARIABLES ABIOTIQUES
- FIGURE 4: HISTOGRAMMES: DISTRIBUTION DES PRATIQUES AGRICOLES
- FIGURE 5: HISTOGRAMME: DISTRIBUTION DE L'ERREUR SUR LE TAUX DE CHANGEMENT DES STOCKS DE CARBONE
- FIGURE 6: BOXPLOTS DES ERREURS DE PREDICTION PAR MODELE DE SIMULATION. DEUX BOXPLOTS AVEC LA MEME LETTRE SONT IDENTIQUES D'APRES LE TEST. LE NOMBRE CORRESPOND AU NOMBRE D'OBSERVATION POUR CHAQUE MODELE.
- FIGURE 7: NUAGE DE POINTS DU TAUX DE STOCKAGE DE CARBONE SIMULE EN FONCTION DU TAUX DE STOCKAGE DE CARBONE MESURE. LA DROITE $Y=X$.
- FIGURE 8 :ERREUR SUR LE TAUX DE CHANGEMENT DES STOCKS DE CARBONE DES MODELES EN FONCTION DE LA TEMPERATURE.
- FIGURE 9: :ERREUR SUR LE TAUX DE CHANGEMENT DES STOCKS DE CARBONE DES MODELES EN FONCTION DU BILAN HYDRIQUE. P-ETP : PRECIPITATION- EVAPOTRANSPIRATION POTENTIELLE (MM). LA DROITE DE REGRESSION ET SON INTERVALLE DE CONFIANCE A 95%
- FIGURE 10 :ERREUR SUR LE TAUX DE CHANGEMENT DES STOCKS DE CARBONE DES MODELES EN FONCTION DE L'ARGILE (G. KG-1). LA DROITE DE REGRESSION ET SON INTERVALLE DE CONFIANCE A 95%
- FIGURE 11: ERREUR SUR LE TAUX DE CHANGEMENT DES STOCKS DE CARBONE DES MODELES EN FONCTION DE CaCO_3 (G. KG-1). LA DROITE DE REGRESSION ET SON INTERVALLE DE CONFIANCE A 95%
- FIGURE 12: :ERREUR SUR LE TAUX DE CHANGEMENT DES STOCKS DE CARBONE DES MODELES EN FONCTION DU PH. LA DROITE DE REGRESSION ET SON INTERVALLE DE CONFIANCE A 95%
- FIGURE 13: :ERREUR SUR LE TAUX DE CHANGEMENT DES STOCKS DE CARBONE DES MODELES EN FONCTION DU STOCK INITIAL DE CARBONE. LA DROITE DE REGRESSION ET SON INTERVALLE DE CONFIANCE A 95%
- FIGURE 14: :ERREUR SUR LE TAUX DE CHANGEMENT DES STOCKS DE CARBONE DES MODELES EN FONCTION DE L'APPORT DE FERTILISANTS ORGANIQUES. LA DROITE DE REGRESSION ET SON INTERVALLE DE CONFIANCE A 95%
- FIGURE 15: :ERREUR SUR LE TAUX DE CHANGEMENT DES STOCKS DE CARBONE DES MODELES EN FONCTION DU TRAVAIL DU SOL. LA DROITE DE REGRESSION ET SON INTERVALLE DE CONFIANCE A 95%
- FIGURE 16: :ERREUR SUR LE TAUX DE CHANGEMENT DES STOCKS DE CARBONE DES MODELES EN FONCTION DES RESIDUS DES CULTURES LA DROITE DE REGRESSION ET SON INTERVALLE DE CONFIANCE A 95%
- FIGURE 17:ERREUR DES MODELES EN FONCTION DES CULTURES INTERMEDIAIRES. LA DROITE DE REGRESSION ET SON INTERVALLE DE CONFIANCE A 95%
- FIGURE 18::ERREUR DES MODELES EN FONCTION DES CULTURES DE COLZA. LA DROITE DE REGRESSION ET SON INTERVALLE DE CONFIANCE A 95%
- FIGURE 19: HISTOGRAMME D'ÉBOULIS DES VALEURS PROPRES
- FIGURE 20: CERCLE DE CORRELATION DE L'ACP AVEC LES ERREURS DES MODELES EN SUPPLEMENTAIRE ET PLAN DES INDIVIDUS (TRAITEMENTS). SEULES LES VARIABLES DE CONTRIBUTION SUPERIEURE A 0.4 SONT REPRESENTÉES.
- FIGURE 21: HISTOGRAMME D'ÉBOULIS DES VALEURS PROPRES
- FIGURE 22: CERCLE DE CORRELATION DE L'ACP AVEC LES ERREURS DES MODELES INCLUSES ET PLAN DES INDIVIDUS (TRAITEMENTS)

7. Liste des tableaux:

- TABEAU 1:TABLEAU DES EFFECTIFS DE LA BASE DE DONNEE
- TABEAU 2: STATISTIQUES DESCRIPTIVES DES VARIABLES ABIOTIQUES
- TABEAU 3:KURTOSIS ET SKEWNESS DES VARIABLES ABIOTIQUES
- TABEAU 4:STATISTIQUES DESCRIPTIVES DES PRATIQUES CULTURALES
- TABEAU 5:KURTOSIS ET SKEWNESS DES PRATIQUES CULTURALES
- TABEAU 6:STATISTIQUES DESCRIPTIVES DE L'ERREUR DE PREDICTION
- TABEAU 7: TABLEAU DES MOYENNES ET DES ECARTS-TYPES DE L'ERREUR PAR MODELE DE SIMULATION
- TABEAU 8: TEST SUR LA PENTE ET L'ORDONNEE A L'ORIGINE DES MODELES EN FONCTION DE LA TEMPERATURE. LES CASES ROUGES POUR LES P-VALEUR < 5% (TEST SIGNIFICATIF)
- TABEAU 9: TEST SUR LA PENTE ET L'ORDONNEE A L'ORIGINE DES MODELES EN FONCTION DU BILAN HYDRIQUE. LES CASES ROUGES POUR LES P-VALEUR < 5% (TEST SIGNIFICATIF)

TABEAU 10: TEST SUR LA PENTE ET L'ORDONNEE A L'ORIGINE DES MODELES EN FONCTION DE L'ARGILE. LES CASES ROUGES POUR LES P-VALEUR < 5% (TEST SIGNIFICATIF)

TABEAU 11: TEST SUR LA PENTE ET L'ORDONNEE A L'ORIGINE DES MODELES EN FONCTION DU CALCAIRE. LES CASES ROUGES POUR LES P-VALEUR < 5% (TEST SIGNIFICATIF)

TABEAU 12: TEST SUR LA PENTE ET L'ORDONNEE A L'ORIGINE DES MODELES EN FONCTION DU PH. LES CASES ROUGES POUR LES P-VALEUR < 5% (TEST SIGNIFICATIF)

TABEAU 13: TEST SUR LA PENTE ET L'ORDONNEE A L'ORIGINE DES MODELES EN FONCTION DU STOCK INITIAL DE CARBONE. LES CASES ROUGES POUR LES P-VALEUR < 5% (TEST SIGNIFICATIF)

TABEAU 14: TEST SUR LA PENTE ET L'ORDONNEE A L'ORIGINE DES MODELES EN FONCTION DE L'APPORT DES FERTILISANTS ORGANIQUES. LES CASES ROUGES POUR LES P-VALEUR < 5% (TEST SIGNIFICATIF)

TABEAU 15: TEST SUR LA PENTE ET L'ORDONNEE A L'ORIGINE DES MODELES EN FONCTION DU TRAVAIL DES SOLS. LES CASES ROUGES POUR LES P-VALEUR < 5% (TEST SIGNIFICATIF)

TABEAU 16: TEST SUR LA PENTE ET L'ORDONNEE A L'ORIGINE DES MODELES EN FONCTION DES RESIDUS ANIMAUX ET VEGETAUX. LES CASES ROUGES POUR LES P-VALEUR < 5% (TEST SIGNIFICATIF)

TABEAU 17: TEST SUR LA PENTE ET L'ORDONNEE A L'ORIGINE DES MODELES EN FONCTION DES CULTURES INTERMEDIAIRES. LES CASES ROUGES POUR LES P-VALEUR < 5% (TEST SIGNIFICATIF)

TABEAU 18: TEST SUR LA PENTE ET L'ORDONNEE A L'ORIGINE DES MODELES EN FONCTION DES CULTURE DE COLZA. LES CASES ROUGES POUR LES P-VALEUR < 5% (TEST SIGNIFICATIF)

TABEAU 19: TABLEAU DES \cos^2 POUR LES DEUX PREMIERES DIMENSIONS.

TABEAU 20: TABLEAU DES \cos^2 POUR LES DEUX PREMIERES DIMENSIONS.

8. Annexes:

8.1. Annexe1: Description des modèles mécanistes de simulation.

8.1.1. Modèle AMG : (Andriulo, Mary, Guérif)

Mis au point dès 1999 par l'INRA de Laon (Andriulo et al, 1999), il est simple d'utilisation et robuste. Il a été adapté à partir du premier modèle de bilan humique réalisé en France par Henin et Dupuis (Hénin et al., 1945). Son pas de temps est l'année. Le grand avantage de ce modèle est qu'il nécessite que peu de paramètres en entrée : un coefficient iso-humique K_1 , qui correspond à la quantité d'humus que l'on peut former à partir des résidus organiques et K_2 traduit la perte par minéralisation du carbone stocké dans l'humus (dépendant des conditions pédoclimatiques) (Figure 23).

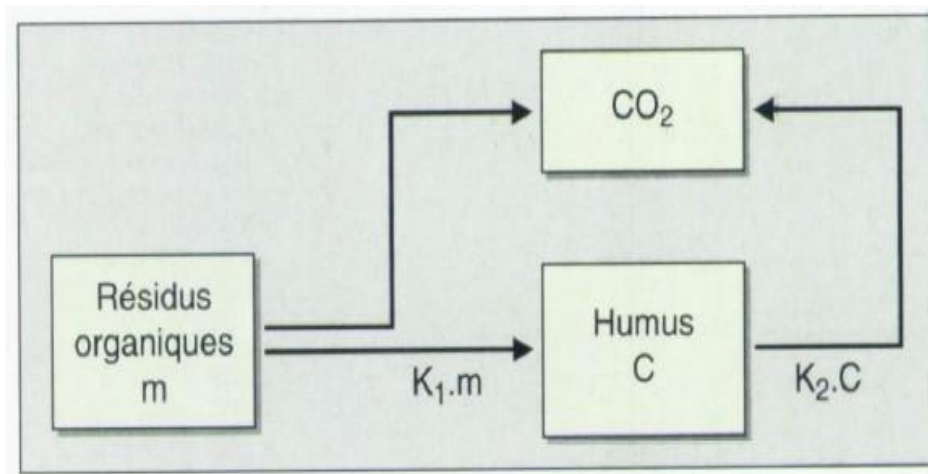


Figure 23: Compartiments du modèle Henin-Dupuis

L'humus est la couche supérieure du sol créée, entretenue et modifiée par la décomposition de la matière organique, principalement par l'action combinée des animaux, des bactéries et des champignons du sol.

8.1.2. Modèle CENTURY :

Contrairement au modèle AMG, le modèle CENTURY est plus vaste et compliqué. Il est à ce jour le meilleur moyen qui permet la compréhension de la bio-géochimie du carbone, de l'azote, du phosphore et du soufre réunis dans un seul modèle. Il est très complet pour le cycle du carbone mais il ne permet pas de faire des prédictions dans les années à venir et demande de nombreuses variables en entrée (Parton et al., 1988 ; Tornquist et al, 2009).

Le modèle "Unforced" consiste à ne pas forcer les entrées de carbone dans le sol issus des plantes.

8.1.3. Modèle RothC

Le modèle RothC permet de calculer la quantité totale de carbone des sols, la quantité de carbone contenue dans la biomasse microbienne et la quantité du carbone 14 des sols. Ce modèle est inspiré des mêmes travaux que pour le modèle Century (Jenkinson, 1990) et propose donc de séparer le carbone des sols en cinq compartiments : La matière organique labile, la matière organique résistante, la matière organique humifiée, la biomasse microbienne et la matière organique inerte (Figure 24).

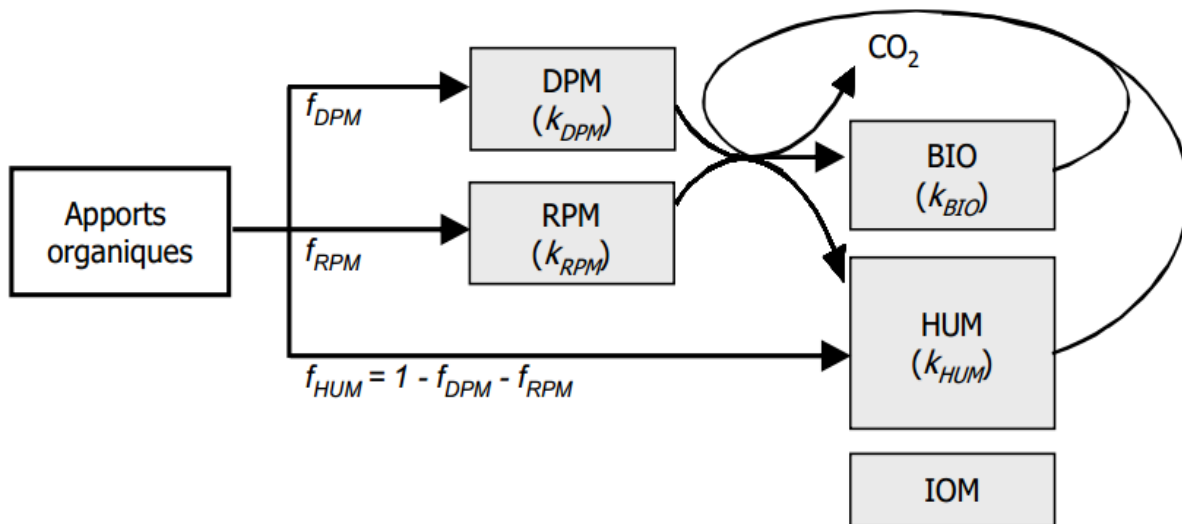


Figure24: Les différents compartiments du modèle RothC

DPM : MO labile	BIO : Biomasse microbienne	
RPM : MO résistante	HUM : MO humifiée	IOM : MO inerte

Le modèle "Relaxed" consiste, comme pour Century, d'initialiser la simulation par une mesure.

8.1.4. Modèle IPCC Tier 1

Le modèle IPCC (En français GIEC : Groupe Intergouvernemental sur l'Evolution du Climat) comporte trois méthodes (Tier1, Tier2 et Tier3). Nous nous intéressons à la méthode Tier 1 qui repose sur des lignes directrices qui permettent de quantifier l'effet des pratiques culturales sur la variation des stocks de COS. Ces lignes directrices sont mises en œuvre sur la période 1990-2010 dans une perspective de comptabilisation des émissions de CO₂ en comptabilité avec les inventaires actuels, et sur des scénarios d'évolution des pratiques à l'horizon 2030 à des fins de prospective et d'appui aux politiques publiques. Afin d'adapter la méthode au contexte français, la base de donnée du Réseau de Mesures de la Qualité des Sols (RMQS), obtenue à l'échelle du territoire, est utilisée pour définir localement des stocks de COS de référence représentatifs.

8.2. Annexe2: Codes R

```
9. #####BOUNOUA El-Mehdi
10. ##### script de Stage M1
11. ##### Sujet: Analyse de données issues de modèles d'évolution du ca
    rbone
12. ##### durée: 15/05/2017 à 15/09/2017
13. ##### Packages----
14. library(tidyr)
15. library(plyr)
16. library(dplyr)
17. library(ggplot2)
18. library(factoextra) ## permet de tracer les graphes dAcp, utilise
    en entrée une PCA de la lib FactoMineR
19. library(FactoMineR)
20. library(e1071) ## kurtosis
21. library(MASS)
22. library(multcomp)
23. library(multcompView)
24.
25. ## Chargement des données des stocks de carbone ----
26. # le tableau des simulations et des observations
27. Predrate <- read.table("D:/embounoua/Principal/Données C/pratiques_
    SOCInitial_SOCChangeRate_ParTraitement.csv",
28.                        na.strings = T, head= T, sep=";")
29.
30. #####chargement du tableau des variables abiotiques----
31. var_abio <- read.table("D:/embounoua/Principal/Données C/entree_ita
    _AMG_19dec2016(IC-Var).csv",
32.                        na.strings = T, head= T, sep=";")
33.
34. var_abio <- var_abio[,c(2,4,5,7,8,9,10)]
35.
36. #####moyenne des données abiotiques
37. Moy_abio <- data.frame(Temperature=tapply(var_abio$Temperature,var_
    abio$ID.Traitement,mean),
38.                        P.ETP=tapply(var_abio$P.ETP,var_abio$ID.Trai
    tement,mean),
39.                        Argile=tapply(var_abio$Argile,var_abio$ID.Tr
    aitement,mean),
40.                        CaCO3=tapply(var_abio$CaCO3,var_abio$ID.Trai
    tement,mean),
41.                        pH=tapply(var_abio$pH,var_abio$ID.Traitement
    ,mean),
42.                        densite.a=tapply(var_abio$Densite.apparente,
    var_abio$ID.Traitement,mean))
43. #####jointure entre Tableaux des erreurs et des var abiotiques----
44. Moy_abio$Traitement <- rownames(Moy_abio)
45.
46. Tab1 <- merge(Moy_abio,Predrate,by.x="Traitement",by.y="id_profil_a
    ial")
47. acprate <- Tab1[,c(14,1,2,3,4,5,6,7,8,9,10,11,12,15,16,19)]
48. Tab1[330,9] <- 1 ## le traitement Fagniere (FAG_L9 est erroné)
49. Tab1[331,9] <- 1
50. Tab1[332,9] <- 1
51. Tab1[333,9] <- 1
52. Tab1[334,9] <- 1
53. Tab1[335,9] <- 1
54. Tab1[333,9] <- 1
55.
```

```

56. ##### séparer le tableau en sous tableaux pour chaque modèle et fair
   e des moyennes de l'erreur par model #####
57. df_AMG<-subset(acprate,model=="AMG")
58. Amg_error <- data.frame(Traitement=unique(df_AMG$Traitement),AMG=ta
   pply(df_AMG$delta_rate_error,df_AMG$Traitement,mean))
59.
60. df_Century<-subset(acprate,model=="Century")
61. Cent_error <- data.frame(Traitement=unique(df_Century$Traitement),C
   ENTURY=tapply(df_Century$delta_rate_error,df_Century$Traitement,mean))
62.
63. df_CenturyUnforced<-subset(acprate,model=="CenturyUnforced")
64. CenturyUnforced_error <- data.frame(Traitement=unique(df_CenturyUnf
   orced$Traitement),CenturyUnforced=tapply(df_CenturyUnforced$delta_rate_e
   rror,df_CenturyUnforced$Traitement,mean))
65.
66. df_IPCCTier1<-subset(acprate, model=="IPCCTier1")
67. IPCCTier1_error <- data.frame(Traitement=unique(df_IPCCTier1$Traite
   ment), IPCCTier1=tapply(df_IPCCTier1$delta_rate_error, df_IPCCTier1$Trai
   tement,mean))
68.
69. df_ORCHIDEE<-subset(acprate,model=="ORCHIDEE")
70. ORCHIDEE_error <- data.frame(Traitement=unique(df_ORCHIDEE$Traiteme
   nt),ORCHIDEE=tapply(df_ORCHIDEE$delta_rate_error,df_ORCHIDEE$Traitement,
   mean))
71.
72. df_RothC<-subset(acprate,model=="RothC")
73. RothC_error <- data.frame(Traitement=unique(df_RothC$Traitement),Ro
   thC=tapply(df_RothC$delta_rate_error,df_RothC$Traitement,mean))
74.
75. df_RothCRelaxed<-subset(acprate,model=="RothCRelaxed")
76. RothCRelaxed_error <- data.frame(Traitement=unique(df_RothCRelaxed$
   Traitement),RothCRelaxed=tapply(df_RothCRelaxed$delta_rate_error,df_Roth
   CRelaxed$Traitement,mean))
77.
78. #####Jointure des sous tableaux----
79. T1 <- merge(Amg_error,Cent_error,by="Traitement")
80. T1 <- merge(T1,CenturyUnforced_error,by="Traitement")
81. T1 <- merge(T1,IPCCTier1_error,by="Traitement")
82. T1 <- merge(T1,ORCHIDEE_error,by="Traitement")
83. T1 <- merge(T1,RothC_error,by="Traitement")
84. T1 <- merge(T1,RothCRelaxed_error,by="Traitement")
85.
86.
87. tab_acprate <- Tab1[,c(1:12,15)] %>%
88.   unique() %>%
89.   merge(T1,.,by="Traitement")
90. tab_acprate <- tab_acprate[-c(5,44),]
91. rownames(tab_acprate) <- tab_acprate$Traitement
92. ##### Ajout de colonne Site----
93. Tab1$Site <-substring(Tab1$Traitement,0,3)
94. grepl("CM", Tab1$Site)
95. Tab1$Site[grepl("CM", Tab1$Site)] <- "CM"
96. table(Tab1$Site)
97.
98. tab_acprate$Site <-substring(tab_acprate$Traitement,0,3)
99. grepl("CM", tab_acprate$Site)
100. tab_acprate$Site[grepl("CM", tab_acprate$Site)] <- "CM"
101. table(tab_acprate$Site)
102. ##### ACP avec tous les modèles en supplémentaire #####
103. acpl <- PCA(tab_acprate[, -c(14,21)], scale.unit=T, quanti.sup=2:8, qu
   ali.sup=1, graph=F)
104. # Graphique des individus

```

```

105.   fviz_pca_ind(acp1, axes = c(1, 2), geom = c("point", "text"),
106.           label = "ind", invisible = "none", labelsizesize = 1,
107.           pointsizesize = 2, habillage = "none",
108.           addEllipses = F, ellipse.level = 0.95,
109.           col.ind = "black", col.ind.sup = c("red", "blue"),
110.           select.ind = list(cos2=0.4))
111.   # Biplot des individus et des variables
112.   fviz_pca_biplot(acp1, axes = c(1, 2), geom = c("point", "text"),
113.           label = "all", invisible = "none", labelsizesize = 2, p
114.           ointsizesize = 0.5,
115.           habillage = "none", addEllipses = FALSE, ellipse.le
116.           vel = 0.95,
117.           col.ind = "black", col.ind.sup = "blue", alpha.ind
118.           = 1,
119.           col.var = "steelblue", alpha.var = 1, col.quantit
120.           p = "blue",
121.           select.var = list(cos2=0.4))
122.   inertiel<-acp1$eig$eigenvalue/sum(acp1$eig$eigenvalue)*100
123.   barplot(inertiel,ylab="% d'inertie")
124.   title("Eboulis des valeurs propres en %")
125.   ##### ACP incluant l'erreur des modèles dans le calcul----
126.   acp2 <- PCA(tab_acprate[, -c(14,21)], scale.unit=T,quali.sup=1,graph
127.           =F)
128.   # Biplot des individus et des variables
129.   fviz_pca_biplot(acp2, axes = c(1, 2), geom = c("point", "text"),
130.           label = "all", invisible = "none", labelsizesize = 2, p
131.           ointsizesize = 0.5,
132.           habillage = "none", addEllipses = FALSE, ellipse.le
133.           vel = 0.95,
134.           col.ind = "black", col.ind.sup = "blue", alpha.ind
135.           = 1,
136.           col.var = "steelblue", alpha.var = 1, col.quantit
137.           p = "blue",
138.           select.var = list(cos2=0.4))
139.   inertie2<-acp2$eig$eigenvalue/sum(acp2$eig$eigenvalue)*100
140.   barplot(inertie2,ylab="% d'inertie",names.arg=round(inertie2,2))
141.   title("Eboulis des valeurs propres en %")
142.   #####Couleurs----
143.   colgeo <- c("grey", "lightskyblue1", "goldenrod2", "yellow", "tan3",
144.           "mediumpurple1", "black", "red", "darkmagenta",
145.           "darkolivegreen", "pink", "burlywood1", "goldenrod4",
146.           "darkolivegreen1", "green", "darkblue",
147.           "lightblue2", "olivedrab4", "skyblue2")

```

```

139.   col_occ <- c("lightskyblue1","lightskyblue1","lightskyblue1","mediu
      mpurple1",
140.               "mediumpurple1","mediumpurple1","red","red",
141.               "red","green","green","green","black",
142.               "black","black",
143.               "pink","pink","pink","brown","brown","brown")
144.   col_symb <- c(4,19,17,4,19,17,4,19,17,4,19,17,4,19,17,4,19,17,4,19,
      17)
145.   #####Temperature----
146.   p <- ggplot(data=Tab1,aes(x=Temperature,y=delta_rate_error,
147.                             col=Site,group=model,shape=Site))
148.
149.   p <- p+ geom_point(size=1.5)+
150.     scale_color_manual(values=col_occ)+
151.     scale_shape_manual(values=col_symb)+
152.     geom_smooth(method="lm",show.legend=F)+
153.     xlab("Temperature °C")+ ylab(expression(paste("Erreur sur le taux
      de changement des stocks de carbone", " ", t.ha^-1)))+
154.     #ggtitle("Erreur des modeles en fonction de la temperature")+
155.     theme_bw()+
156.     facet_wrap(~ model)
157.   p
158.   ggsave("rateTemperature.jpeg")
159.
160.   ##### pH----
161.   p <- ggplot(data=Tab1,aes(x=pH,y=delta_rate_error,
162.                             col=Site,group=model,shape=Site))
163.   p <- p+ geom_point(size=1.5)+
164.     scale_color_manual(values=col_occ)+
165.     scale_shape_manual(values=col_symb)+
166.     scale_size_manual(values=rep(1,21))+
167.     geom_smooth(method="lm",show.legend=F)+
168.     xlab("pH")+ ylab(expression(paste("Erreur sur le taux de changeme
      nt des stocks de carbone", " ", t.ha^-1)))+
169.     # ggtitle("Erreur des modèles en fonction du pH")+
170.     theme_bw()+
171.     facet_wrap(~ model)
172.   p
173.   ggsave("ratepH.jpeg")

```

```

174. ##### P-ETP----
175. p <- ggplot(data=Tab1,aes(x=P.ETP,y=delta_rate_error,
176.                           col=Site,group=model,shape=Site))
177. p <- p+ geom_point(size=1.5)+
178.   scale_color_manual(values=col_occ)+
179.   scale_shape_manual(values=col_symb)+
180.   geom_smooth(method="lm",show.legend=F)+
181.   xlab("P-ETP")+ylab(expression(paste("Erreur sur le taux de change
      ment des stocks de carbone", " ", t.ha^-1)))+
182.   # ggtitle("Erreur des modeles en fonction de P-ETP")+
183.   theme_bw()+
184.   facet_wrap(~ model)
185. p
186. ggsave("rateP-ETP.jpeg")
187. #####Argile----
188. p <- ggplot(data=Tab1,aes(x=Argile,y=delta_rate_error,
189.                           col=Site,group=model,shape=Site))
190. p <- p+ geom_point(size=1.5)+
191.   scale_color_manual(values=col_occ)+
192.   scale_shape_manual(values=col_symb)+
193.   geom_smooth(method="lm",show.legend=F)+
194.   xlab(" Argile")+ ylab(expression(paste("Erreur sur le taux de cha
      ngement des stocks de carbone", " ", t.ha^-1)))+
195.   theme_bw()+
196.   facet_wrap(~ model)
197. p
198. ggsave("rateArgile.jpeg")
199. #####Calcaire----
200. p <- ggplot(data=Tab1,aes(x=CaCO3,y=delta_rate_error,
201.                           col=Site,group=model,shape=Site))
202. p <- p+ geom_point(size=1.5)+
203.   scale_color_manual(values=col_occ)+
204.   scale_shape_manual(values=col_symb)+
205.   geom_smooth(method="lm",show.legend=F)+
206.   xlab("CaCO3")+ ylab(expression(paste("Erreur sur le taux de chang
      ement des stocks de carbone", " ", t.ha^-1)))+
207.   theme_bw()+
208.   facet_wrap(~ model)
209. p

```

```

210. ggsave("rateCalcaire.jpeg")
211. #####stock initial de carbone----
212. p <- ggplot(data=Tab1,aes(x=soc_initial,y=delta_rate_error,
213.                             col=Site,group=model,shape=Site))
214. p <- p+ geom_point(size=1.5)+
215.     scale_color_manual(values=col_occ)+
216.     scale_shape_manual(values=col_symb)+
217.     geom_smooth(method="lm",show.legend=F)+
218.     xlab("Stock initial de COS")+ ylab(expression(paste("Erreur sur l
e taux de changement des stocks de carbone", " ", t.ha^-1)))+
219.     theme_bw()+
220.     facet_wrap(~ model)
221. p
222. ggsave("ratesoc_initial.jpeg")
223.
224. ##### Fertilisant organique ----
225. p <- ggplot(data=Tab1,aes(x=omad_avg,y=delta_rate_error,
226.                             col=Site,group=model,shape=Site))
227. p <- p+ geom_point(size=1.5)+
228.     scale_color_manual(values=col_occ)+
229.     scale_shape_manual(values=col_symb)+
230.     geom_smooth(method="lm",show.legend=F)+
231.     xlab("fertOrg")+ ylab(expression(paste("Erreur sur le taux de cha
ngement des stocks de carbone", " ", t.ha^-1)))+
232.     theme_bw()+
233.     facet_wrap(~ model)
234. p
235. ggsave("ratefertOrg.jpeg")
236. ##### labour----
237. p <- ggplot(data=Tab1,aes(x=labour_avg,y=delta_rate_error,
238.                             col=Site,group=model,shape=Site))
239. p <- p+ geom_point(size=1.5)+
240.     scale_color_manual(values=col_occ)+
241.     scale_shape_manual(values=col_symb)+
242.     geom_smooth(method="lm",show.legend=F)+
243.     xlab("labour")+ ylab(expression(paste("Erreur sur le taux de chan
gement des stocks de carbone", " ", t.ha^-1)))+
244.     theme_bw()+
245.     facet_wrap(~ model)

```

```

246. p
247. ggsave("ratelabour.jpeg")
248. #####résidus----
249. p <- ggplot(data=Tab1,aes(x=residus_avg,y=delta_rate_error,
250.                           col=Site,group=model,shape=Site))
251. p <- p+ geom_point(size=1.5)+
252.   scale_color_manual(values=col_occ)+
253.   scale_shape_manual(values=col_symb)+
254.   geom_smooth(method="lm",show.legend=F)+
255.   xlab("résidus")+ ylab(expression(paste("Erreur sur le taux de cha
ngement des stocks de carbone", " ", t.ha^-1)))+
256.   theme_bw()+
257.   facet_wrap(~ model)
258. p
259. ggsave("rateResidus.jpeg")
260. #####culture intérieure----
261. p <- ggplot(data=Tab1,aes(x=cultinter_avg,y=delta_rate_error,
262.                           col=Site,group=model,shape=Site))
263. p <- p+ geom_point(size=1.5)+
264.   scale_color_manual(values=col_occ)+
265.   scale_shape_manual(values=col_symb)+
266.   geom_smooth(method="lm",show.legend=F)+
267.   xlab("culture intermédiaire")+ ylab(expression(paste("Erreur sur
le taux de changement des stocks de carbone", " ", t.ha^-1)))+
268.   theme_bw()+
269.   facet_wrap(~ model)
270. p
271. ggsave("ratecultinter.jpeg")
272. #####culture de colza----
273. p <- ggplot(data=Tab1,aes(x=cultosrlike,y=delta_rate_error,
274.                           col=Site,group=model,shape=Site))
275. p <- p+ geom_point(size=1.5)+
276.   scale_color_manual(values=col_occ)+
277.   scale_shape_manual(values=col_symb)+
278.   geom_smooth(method="lm",show.legend=F)+
279.   xlab("culture de colza")+ ylab(expression(paste("Erreur sur le ta
ux de changement des stocks de carbone", " ", t.ha^-1)))+
280.   theme_bw()+
281.   facet_wrap(~ model)

```

```

282. p
283. ggsave("ratecultosrlike.jpeg")
284. ##### représentation du carbone simulé / mesuré #####
285.
286. p <- ggplot(data=Tab1,aes(x=delta_rate_sim,y=delta_rate_obs,col=Site,
287.                             group=model,shape=Site))
288.
289. p <- p+ geom_point(size=2)+
290.   xlab("C mesure")+ ylab("C simule")+
291.   scale_color_manual(values=col_occ)+
292.   scale_shape_manual(values=col_symb)+
293.   ggtitle("Simulation/mesure")+
294.   theme_bw()+
295.   facet_wrap(~model)
296. p <- p + geom_abline(intercept=0, slope=1)
297. p
298. ggsave("rate_sim_mesure.jpeg")
299. ##### les coeff de régression tous les modèles TEMPERATURE----
300. summary(lm(tab_acprate$AMG ~ tab_acprate$Temperature))
301. summary(lm(tab_acprate$CENTURY ~ tab_acprate$Temperature))
302. summary(lm(tab_acprate$CenturyUnforced ~ tab_acprate$Temperature))
303. summary(lm(tab_acprate$IPCCTier1 ~ tab_acprate$Temperature))
304. summary(lm(tab_acprate$ORCHIDEE ~ tab_acprate$Temperature))
305. summary(lm(tab_acprate$RothC ~ tab_acprate$Temperature))
306. summary(lm(tab_acprate$RothCRelaxed ~ tab_acprate$Temperature))
307. ##### les coeff de régression tous les modèles P-ETP----
308. summary(lm(tab_acprate$AMG ~ tab_acprate$P.ETP))
309. summary(lm(tab_acprate$CENTURY ~ tab_acprate$P.ETP))
310. summary(lm(tab_acprate$CenturyUnforced ~ tab_acprate$P.ETP))
311. summary(lm(tab_acprate$IPCCTier1 ~ tab_acprate$P.ETP))
312. summary(lm(tab_acprate$ORCHIDEE ~ tab_acprate$P.ETP))
313. summary(lm(tab_acprate$RothC ~ tab_acprate$P.ETP))
314. summary(lm(tab_acprate$RothCRelaxed ~ tab_acprate$P.ETP))
315. #####histogramme et stats desc du taux d'erreur----
316. hist(Tab1$delta_rate_error,col="blue",main="Histogramme d'erreur sur
    le taux de changement des stocks de carbone",

```

```

317.       xlab=expression(paste("Erreur", " ", t.ha^-1)),ylab="Fréquence
    ")

```

```

318. summary(Tab1$delta_rate_error)
319. kurtosis(Tab1$delta_rate_error)
320. skewness(Tab1$delta_rate_error)
321.
322. ##### Test de Kruskal Wallis #####
323. #####fonction qui affiche le resultat du test de kruskal-wallis sur
    le boxplot ----
324. #' @title generate_label_df
325. #'
326. #' @description Permet de créer les labels d'un test statistiques (
    ici wilcoxon, à adapter pour d'autre test) + nbr analyse par groupe

```



```

327.   #' @param lev : nom des groupes dont on souhaite tester la différen
      ce significative
328.   #' @param db : nom de la base de données
329.   #' @param value : nom du champs contenant les valeurs à évaluer
330.   #' @param letter : option pour les lettres (letters, LETTERS)
331.
332.
333.   tri.to.squ<-function(x)
334.   {
335.     rn<-row.names(x)
336.     cn<-colnames(x)
337.     an<-unique(c(cn,rn))
338.     myval<-x[!is.na(x)]
339.     mymat<-matrix(1,nrow=length(an),ncol=length(an),dimnames=list(an,
      an))
340.     for(ext in 1:length(cn))
341.     {
342.       for(int in 1:length(rn))
343.       {
344.         if(is.na(x[row.names(x)==rn[int],colnames(x)==cn[ext]])) next
345.         mymat[row.names(mymat)==rn[int],colnames(mymat)==cn[ext]]<-x[
      row.names(x)==rn[int],colnames(x)==cn[ext]]
346.         mymat[row.names(mymat)==cn[ext],colnames(mymat)==rn[int]]<-x[
      row.names(x)==rn[int],colnames(x)==cn[ext]]
347.       }
348.
349.     }
350.     return(mymat)
351.   }
352.
353.   generate_label_df <- function(
354.     db,
355.     value,
356.     lev,
357.     letter,
358.     position
359.   )
360.   {
361.     wilcotest <- pairwise.wilcox.test(db[,value], db[,lev])$p.value
362.     mymat <- tri.to.squ(wilcotest)
363.     myletters <- multcompLetters(mymat,compare("<="",threshold=0.05,Le
      tters=letters)
364.     plot.labels <- names(myletters[['Letters']])
365.
366.     #changement $value par [value]
367.     boxplot.df <- ddply(db, lev, function (x) quantile(x[value],posit
      ion,na.rm=TRUE) + 0.2)
368.     colnames(boxplot.df) <- c("lev","V1")
369.
370.     nbranalyse.df <- ddply(db, lev, function (x) nrow(x[value]))
371.     colnames(nbranalyse.df) <- c("lev","Nbr_analyse")
372.
373.     # Create a data frame out of the factor levels and Tukey's homoge
      nous group letters
374.     plot.levels <- data.frame(plot.labels, labels = myletters[['Lette
      rs']],
375.                               stringsAsFactors = FALSE)
376.
377.     # Jointure
378.     labels.df <- merge(plot.levels, boxplot.df, by.x = 'plot.labels',
      by.y = "lev", sort = FALSE)

```

```

379.     labels.df <- merge(labels.df, nbranalyse.df, by.x = 'plot.labels'
, by.y = "lev", sort = FALSE)
380.
381.     return(labels.df)
382.   }
383.
384.
385.   # From http://stackoverflow.com/questions/13649473/add-a-common-leg
end-for-combined-ggplots
386.   grid_arrange_shared_legend <- function(..., nrow = 1, ncol = length
(list(...)), position = c("bottom", "right")) {
387.
388.     plots <- list(...)
389.     position <- match.arg(position)
390.     g <- ggplotGrob(plots[[1]] + theme(legend.position = position))$g
robs
391.     legend <- g[[which(apply(g, function(x) x$name) == "guide-box")]
]
392.     lheight <- sum(legend$height)
393.     lwidth <- sum(legend$width)
394.     gl <- lapply(plots, function(x) x + theme(legend.position = "none
"))
395.     gl <- c(gl, nrow = nrow, ncol = ncol)
396.
397.     combined <- switch(position,
398.                           "bottom" = arrangeGrob(do.call(arrangeGrob, gl
),
399.                                                     legend,
400.                                                     ncol = 1,
401.                                                     heights = unit.c(unit(1
, "npc") - lheight, lheight)),
402.                           "right" = arrangeGrob(do.call(arrangeGrob, gl
),
403.                                                     legend,
404.                                                     ncol = 2,
405.                                                     widths = unit.c(unit(1,
"npc") - lwidth, lwidth)))
406.     grid.newpage()
407.     grid.draw(combined)
408.     return(combined)
409.   }
410.
411.   # Fonction pour ajouter un thème perso sur ggplot2
412.   # Aide !http://ggplot2.tidyverse.org/reference/theme.html
413.   theme_perso <- function(position="bottom",...){
414.     theme_classic()+
415.     theme(
416.       axis.text.x = element_text(size = 11, colour = "black"),#,face
e = "bold"),
417.       axis.text.y = element_text(size = 11, colour = "black"),#,face
e = "bold"),face = "bold"),
418.       axis.line.x = element_line(colour = "black", size = 0.7),
419.       axis.line.y = element_line(colour = "black", size = 0.7),
420.       plot.title = element_text(size = 14, face = "bold"),
421.       text = element_text(size = 12),
422.       strip.background = element_rect(colour = "white", fill = "gre
y"),#pour facet_wrap
423.       strip.text.x = element_text(colour = "black", face = "bold"),
#pour facet_wrap
424.       axis.title = element_text(face="bold"),
425.       legend.position = position)

```

```

426.   }
427.
428.   # Pour afficher la droite de regression + l'équation et le R2
429.   # from https://gist.github.com/ottadini/6882677
430.   # Source: http://stackoverflow.com/q/7549694/857416
431.   lm_eqn = function(m) {
432.     # Displays regression line equation and R^2 value on plot
433.     # Usage:
434.     # p + annotate("text", x=25, y=300, label=lm_eqn(lm(y ~ x, df)),
435.     parse=TRUE)
436.     l <- list(a = format(coef(m)[1], digits = 2),
437.              b = format(abs(coef(m)[2]), digits = 2),
438.              r2 = format(summary(m)$r.squared, digits = 3));
439.
440.     if (coef(m)[2] >= 0) {
441.       eq <- substitute(italic(y) == a + b % italic(x)*", "~italic(r
442.       )^2~"=="~r2,l)
443.     } else {
444.       eq <- substitute(italic(y) == a - b % italic(x)*", "~italic(r
445.       )^2~"=="~r2,l)
446.     }
447.     as.character(as.expression(eq));
448.   }
449.   # le tableau des observations
450.   Entree_ita_AMG <- read.table("D:/embounoua/Principal/Données C/entr
451.   ee_ita_AMG_19dec2016(IC-Var).csv",
452.                               na.strings = T, head= T, sep=";")
453.   var_abio <- read.table("D:/embounoua/Principal/Données C/entree_ita
454.   _AMG_19dec2016(IC-Var).csv",
455.                         na.strings = T, head= T, sep=";")
456.   var_abio <- var_abio[,c(2,4,5,7,8,9,10)]

```

```

455.
456.
457.   krustable <- Predrate %>%
458.     left_join(Entree_ita_AMG, by = c('id_profil_aial' = 'ID.Traitemen
459.     t')) %>%
460.     filter(Stock.de.carbone>0) %>%
461.     dplyr::select(id_profil_aial,model,delta_rate_sim,delta_rate_obs,
462.     delta_rate_error)
463.     ##### couleur des boxplots----
464.     couleur <- c("grey", "lightskyblue1", "goldenrod2", "yellowgreen",
465.     "tan3", "mediumpurple1",
466.     "orange4")
467.     #####noms des axes----
468.     lablex <- "modèle"

```

```

467.   labley <- "Taux de l'erreur (t/ha)"

```

```

468.   #####test de kruskal-wallis----
469.   wilcotest <- generate_label_df(db=krustable,value="delta_rate_error
470.   ",lev="model",letters,position=1)
471.   #####boxplots----
472.   p <- ggplot(krustable)+
473.     geom_boxplot(aes_string(y="delta_rate_error",x="model",fill="mode
474.     l"))+
475.     geom_hline(yintercept = 0,color="red") +
476.     scale_fill_manual(values=couleur,name="test") +

```

```

475.     geom_text(data = wilcotest, vjust=-0.75,
476.             aes(x = plot.labels, y = V1,
477.                 label = paste(labels,"/", Nbr_analyse, sep=""))) +
478.     scale_x_discrete(lablex)+scale_y_continuous(labley)+theme_perso()
+
479.     theme(legend.position="none") # + facet_wrap(~period)
480. p
481. ggsave ("Taux erreur boxplot et test de krus-wallis.jpeg")
482.
483. #####Stat descriptives des var abiotiques----
484. par(mfrow=c(2,3))
485. hist(Tab1$Temperature,main="Temperature",xlab="",ylab="",col="blue"
)
486. hist(Tab1$P.ETP,main="P-ETP", xlab="", ylab="",col="blue")
487. hist(Tab1$Argile,main="Argile", xlab="", ylab="",col="blue")
488. hist(Tab1$CaCO3,main="Calcaire", xlab="",ylab="",col="blue")
489. hist(Tab1$pH,main="pH", xlab="", ylab="",col="blue")
490. hist(Tab1$soc_initial,main="Stock initial de carbone", xlab="",ylab=
="",col="blue")
491. par(mfrow=c(2,2))
492. hist(Tab1$omad_avg,main="Fertilisant organique", xlab="",ylab="",col
="blue")
493. hist(Tab1$labour_avg,main="Labour", xlab="",ylab="",col="blue")
494. hist(Tab1$residus_avg, main="Résidus", xlab="", ylab="",col="blue")
495. hist(Tab1$cultinter_avg,main="Culture intermédiaire", xlab="",ylab=
="",col="blue")

```

496. Références

- Andriulo, A., Mary, B., Guerif, J., 1999. Modelling soil carbon dynamics with various cropping sequences on the rolling pampas. *Agronomie* 19, 365–377. doi:<http://dx.doi.org/10.1051/agro:19990504>.
- Capéraà, P., & Van Cutsem, B. (1988). *Méthodes et modèles en statistique non paramétrique: exposé fondamental* (Vol. 1). Presses Université Laval.
- Davison A.C. (2008), *Statistical Models*, [Cambridge University Press](#)
- Dubitzky, W., Wolkenhauer, O., Yokota, H., & Cho, K. H. (2013). *Encyclopedia of systems biology*. Springer Publishing Company, Incorporated.
- Hénin S., Dupuis M. (1945) Essai de bilan de la matière organique des sols. *Annales agronomiques*, 15 (1), 161-172.
- IPCC Climate Change, 2014. Mitigation of Climate Change (eds Edenhofer, O. R. et al.) Cambridge Univ. Press.
- Jenkinson DS (1990) The turnover of organic carbon and nitrogen in soil. *Philos Trans R Soc Lond B* 329:361–368
- Krinner, G., Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., Ciais, P.; Sitch, S; Prentice, I. C. (2005). A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. *Global Biogeochemical Cycles*, 19(1).
- Lal, R., Follett, R., Stewart, B. and Kimble, J., 2007. Soil carbon sequestration to mitigate climate change and advance food security. *Soil Sci* 172, 943-956.
- Meersmans, J., Martin, M.P., De Ridder, F., Lacarce, E., Wetterlind, J., De Baets, S., Le Bas, C., Louis, B.P., Orton, T.G., Bispo, A. and Arrouays, D., 2012a. A novel soil organic C model using climate, soil type and management data at the national scale in France. *Agron. Sustain. Dev.* 32 (4), 873-888.
- Parton, W. J., Stewart, J. W. B., and Cole, C. V., 1988. Dynamics of C,N, P and S in grassland soils -a model, *Biogeochemistry*, 5, 109-131.
- R Core Team., 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Smith, P., Martino, D., Cai, Z., Gwary, D., Janzen, H., Kumar, P., McCarl, B., Ogle, S., F. O'Mara, Rice, C., Scholes, B., Sirotenko. O., 2007. Agriculture. In *Climate Change: Mitigation. Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* [B. Metz, O.R. Davidson, P.R. Bosch, R.

Dave, L.A. Meyer (eds)], Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007.

TORNQUIST C.G., MIELNICZUK J., CERRI C.E.P.. 2009. Modeling soil organic carbon dynamics in Oxisols of Ibirubá (Brazil) with the Century Model. *Soil and Tillage Research*. Numéro 105, Pages 33-43.

Tosser, V., Eglin, T., Bardy, M., Besson, A., & Martin, M. (2014). Evaluation des stocks de carbone organique des sols cultivés de France. *Etude et Gestion des Sols*, 21, 7-23.

WENQIANG X., XI C., GEPING L., QING L.. 2011. Using the CENTURY model to assess the impact of land reclamation and management practices in oasis agriculture on the dynamics of soil organic carbon in the arid region of North-western China. *Ecological Complexity*. Numéro 8, Pages 30-37.

Zhang, W., Yu, Y. Q., Li, T. T., Sun, W. J. Huang, Y., 2014. Net Greenhouse Gas Balance in China's Croplands over the Last Three Decades and Its Mitigation Potential. *Environ Sci. Technol.*, 2589-2597.