

# **Les déterminants de l'utilisation de Glyphosate en France**

Alexandre Marlet / Lucas Gerardy France / Caleb Colin / Brice Perroud

2025-04-25

## Table des matières

<b>LES DETERMINANTS DE L'UTILISATION DE GLYPHOSATE EN FRANCE .....</b>	<b>1</b>
<b>INTRODUCTION .....</b>	<b>3</b>
SOURCES DES DONNEES .....	3
CHARGEMENT ET FUSION DES BASES DE DONNEES .....	3
<b>DESCRIPTION DES BASES DE DONNEES .....</b>	<b>4</b>
<b>DICTIONNAIRE DES VARIABLES.....</b>	<b>4</b>
<i>Représentation graphique des données et intuitions .....</i>	<i>5</i>
<b>STATISTIQUES DESCRIPTIVES .....</b>	<b>6</b>
STATISTIQUES UNIVARIEES.....	6
STATISTIQUES BIVARIEES .....	8
<i>Correlations entre les variables.....</i>	<i>8</i>
<i>Test de plusieurs modèles .....</i>	<i>9</i>
NOTRE MEILLEUR MODELE A ESTIMER (MCO4).....	10
<i>Meilleur modèle estimé et interprétation des paramètres .....</i>	<i>10</i>
<i>Qualité d'ajustement.....</i>	<i>11</i>
<i>Analyse du R2 .....</i>	<i>11</i>
RESPECT DES HYPOTHESES H1 A H5.....	12
<i>H1 L'estimation de l'erreur est nulle.....</i>	<i>12</i>
<i>H2 La variance de l'erreur est constante (Test d'hétéroscedasticité).....</i>	<i>12</i>
<i>H3 La matrice X est non aléatoire, la variable est non aléatoire pour toutes les observations.....</i>	<i>18</i>
<i>H4 Le modèle est correctement spécifié .....</i>	<i>18</i>
<i>H5 La matrice X est de plein rang.....</i>	<i>18</i>
MODELE FINAL .....	19
<i>Interprétation .....</i>	<i>19</i>
<i>Performance du modèle .....</i>	<i>20</i>
CONCLUSION .....	20

# Introduction

Les pesticides prennent une place importante dans les productions agricoles en France pour faire face à une concurrence internationale et aux espèces parasites envahissantes qui réduisent les rendements des productions agricoles. Face au constat de ce problème, plusieurs substances s'imposent dont celle faisant partie de notre étude : "le Glyphosate". Dans notre travail, nous porterons notre analyse sur cette substance au travers de la problématique suivante :

Quels sont les déterminants de la quantité de Glyphosate achetée dans les départements français ? Nous analyserons dans notre travail l'impact des types de culture dans le département (céréales, vignes), sa surface agricole utilisée et sa production brute standard sur la quantité de glyphosate achetée.

Autrement dit, par l'analyse de la quantité de glyphosate achetée par département et de la culture dominante, la production brute standard et la surface agricole utilisée par département, nous chercherons à expliquer si une culture a un impact significatif sur la quantité de glyphosate achetée et chercherons si cette relation est davantage significative en fonction de la PBS ou de la SAU.

Dans notre étude, nous mobiliserons deux sources principales de données. D'une part, les données issues du recensement agricole de 2020 (dernières données en date), qui fournissent des informations structurelles sur les caractéristiques agricoles des départements (types de cultures, surfaces agricoles, production brute standard). D'autre part, les données des quantités achetées de glyphosate de 2022.

Bien que ces deux jeux de données proviennent de périodes différentes, leur agrégation est économiquement justifiée par la relative stabilité des structures agricoles sur une courte période de deux ans. En effet, les évolutions de l'organisation agricole à l'échelle départementale restent généralement limitées à court terme, rendant les variables de 2020 représentatives de la situation contemporaine en 2022. Toutefois, nous reconnaissons qu'une légère évolution, notamment liée aux politiques environnementales récentes, pourrait avoir affecté certaines pratiques. Cette limite sera prise en compte dans l'interprétation des résultats.

## Sources des données

substance (<https://www.data.gouv.fr/fr/datasets/achats-de-pesticides-par-code-postal/>)

vignes (<https://agreste.agriculture.gouv.fr/agreste-web/disaron/Carte-RA-partvigne20/detail/>)

pbs (<https://agreste.agriculture.gouv.fr/agreste-web/disaron/Carte-RA-pbs20/detail/>)

sau (<https://agreste.agriculture.gouv.fr/agreste-web/disaron/Carte-RA-sau20/detail/>)

cereales (<https://agreste.agriculture.gouv.fr/agreste-web/disaron/Carte-RA-partcereoleopro20/detail/>)

## Chargement et fusion des bases de données

Nous fusionnons les bases de données pour obtenir une seule et même base contenant les informations utiles. Ces bases sont réunies par la variable en commun "code département". Les informations manquantes représentées par le symbole N/A seront comptabilisées comme une valeur nulle c'est-à-dire "0". Dans un premier temps nous avons retiré ces valeurs, mais nous perdions trop d'information et cela faussait l'analyse et les modèles. Nous avons fait le choix de remplacer les N/A par "0" de manière à garder la donnée des autres variables pour les départements ayant N/A comme information et ne pas omettre totalement la ligne sous prétexte que qu'une donnée est manquante. Pour être encore plus précis nous pourrions dans un second temps envisager de faire la moyenne des départements voisins pour déduire la donnée d'un département. Quelque soit la méthode, une donnée manquante ne peut pas être remplacée, elle peut aux mieux être estimée et approchée le tout est de chercher à l'approcher le plus fidèlement et réalistiquement possible et cela requiert une maîtrise du sujet et des données traitées. Par exemple dans notre cas, la culture de vignes est souvent très localisée: les appellations changent parfois d'une parcelle à une autre ce qui détermine la prépondérance de la culture viticole à certains endroits et sont absentes dans d'autres.

Faire la moyenne de trois départements dont deux ayant une appellation prestigieuse (particulièrement pourvu en vigne) et un troisième sans appellation particulière (avec aucune vignes), si la donnée est manquante pour le troisième alors l'estimer en moyenne de ses voisins serait faux. Mais l'estimer par un "0" serait faux aussi. Il s'agit alors de l'estimer en trouvant une méthode adaptée.

## Description des bases de données

Nous travaillons sur 5 bases de données que nous fusionnons pour étudier l'utilisation de substances phytosanitaires par département en fonction des cultures sur des observations de 2022.

## Dictionnaire des variables

code\_departement - numéro indicatif pour chaque département

departement - nom du département

somme\_dep - quantité de glyphosate achetée dans le département (exprimée en Kg)

cereales - part des céréales et oléagineux dans la SAU (en % de la SAU)

vignes - part de vignes dans la SAU (en % de la SAU)

pbs - production brute standard, production potentielle totale des exploitations par département, résultant des valeurs moyennes des rendements et des prix observés sur la période 2015 à 2019, exprimée en euros.

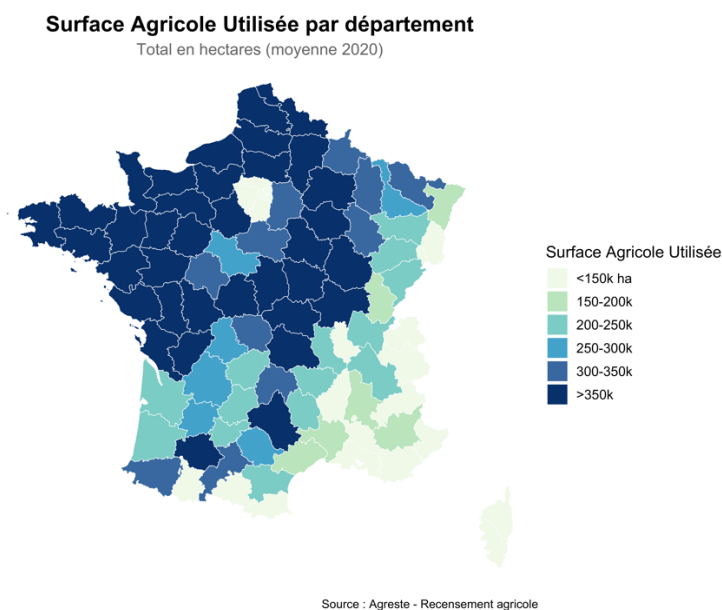
log\_pbs - le logarithme de la variable PBS

sau - superficie agricole utilisée (en hectares), comprenant les céréales, les oléagineux, protéagineux et plantes à fibres, les autres plantes industrielles destinées à la transformation, les cultures fourragères et les surfaces toujours en herbe, les légumes secs et frais, les fraises et les melons, les pommes de terre, les fleurs et plantes ornementales, les vignes, les autres cultures permanentes (vergers, petits fruits, pépinières ligneuses), les jachères, les jardins et vergers familiaux.

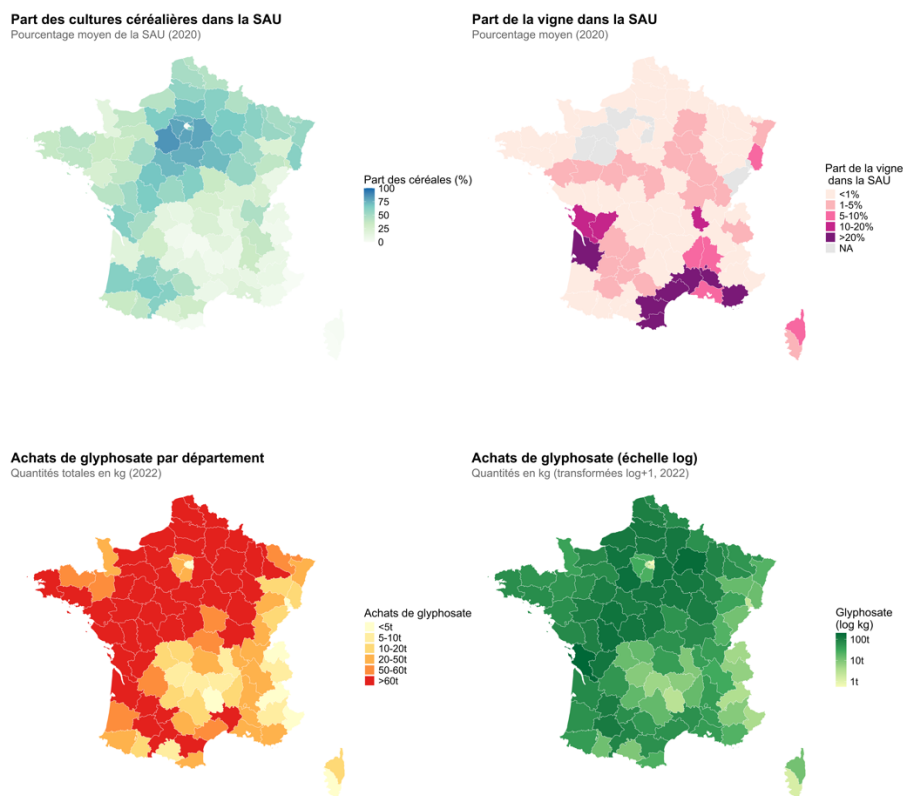
log\_sau - le logarithme de la variable SAU

# Représentation graphique des données et intuitions

## Carte France et répartition



Cette première carte permet de se rendre compte de la répartition des terres agricoles en France. Le Sud-Est, la Corse ainsi que la région parisienne proche sont plutôt faiblement pourvus de terres agricoles. Cela s'explique principalement par le relief pour le Sud-Est et la Corse, ainsi que par la grande zone urbaine de Paris et sa périphérie.











Sources : Agreste - Recensement agricole 2020 | BNVD 2022

On constate que les zones avec une importante production viticole comme la région bordelaise, la côte méditerranéenne ainsi que la Bourgogne sont des zones où l'achat de glyphosate est important. On fait le même constat dans les zones à forte production céréalière (plutôt dans le nord de la France, en périphérie de la région parisienne et dans le Sud-Ouest). A contrario, dans les zones montagneuses, moins propices à ces cultures, représentant davantage des terres d'élevage (grande surface de pâturage), on y retrouve une plus faible quantité de glyphosate acheté comme on le constate sur la carte principalement dans les Alpes, le Massif central et les Pyrénées. Notre intuition nous induit vers une forte corrélation entre la production viticole et l'utilisation de glyphosate, tandis que la corrélation entre la production céréalière et l'utilisation de glyphosate est moins évidente, et peut être même fausse au vu des graphiques. Tout de même, si nous ne pouvons pas clairement faire un lien direct entre les cultures céréalières et une forte utilisation de glyphosate, force est de constater une évidente corrélation des régions avec les plus fortes surfaces agricoles utilisées (SAU) et les régions avec les plus fortes quantités de glyphosate achetées.

Les données manquantes concernant la part de vignes dans la SAU sont dues au fait qu'il n'y ait pas de recensement agricole pour certaines régions puisqu'il n'y a pas ou peu d'activité agricole, c'est le cas de la Seine-Saint-Denis.

## Statistiques descriptives

### Statistiques univariées

	Unique	Missing Pct.	Mean	SD	Min	Median	Max	Histogram
somme_dep	101	0	57281.7	48425.4	1.8	51033.3	199723.1	
cereales	93	0	36.5	25.0	0.0	36.6	84.9	
vignes	40	0	4.0	10.1	0.0	0.2	50.1	
pbs	101	0	645787.6	517305.2	103.0	518664.0	2685717.0	
sau	101	0	266144.4	149930.1	1.0	284438.0	557179.0	
log_pbs	101	0	13.6	1.6	5.3	13.9	15.5	
log_sau	101	0	12.7	2.0	0.9	13.3	13.9	
log_somme_dep	101	0	10.1	2.1	0.6	10.8	12.2	

On peut déduire de ce tableau que la quantité moyenne de Glyphosate achetée par département en France est de 57281.7 kg. Tandis que le département achetant le minimum de Glyphosate en achète 1,8 kg, et le département qui en achetant le plus en utilise 199 723,1 kg. Les observations comptées pour céréales sont de 93 car ici seulement les observations uniques (comprendre différentes) sont prises en compte pour vignes c'est la même chose. On en conclut que beaucoup d'observations sont identiques (les valeurs sont similaires dans plusieurs départements).

TOP 3 DES DÉPARTEMENTS LES PLUS UTILISATEURS DE GLYPHOSATE	
Département	Quantité (kg)
Charente-Maritime	199 723.1
Marne	175 663.8
Eure-et-Loir	155 414.4

Source : Données BNVD 2022 - Ministère de l'Agriculture

TOP 3 des départements céréaliers		
Département	Part de céréales (%)	SAU totale
Eure-et-Loir	84.9	446 392
Essonne	79.8	83 078
Yvelines	79.4	89 291

Source : Agreste - Recensement agricole 2020

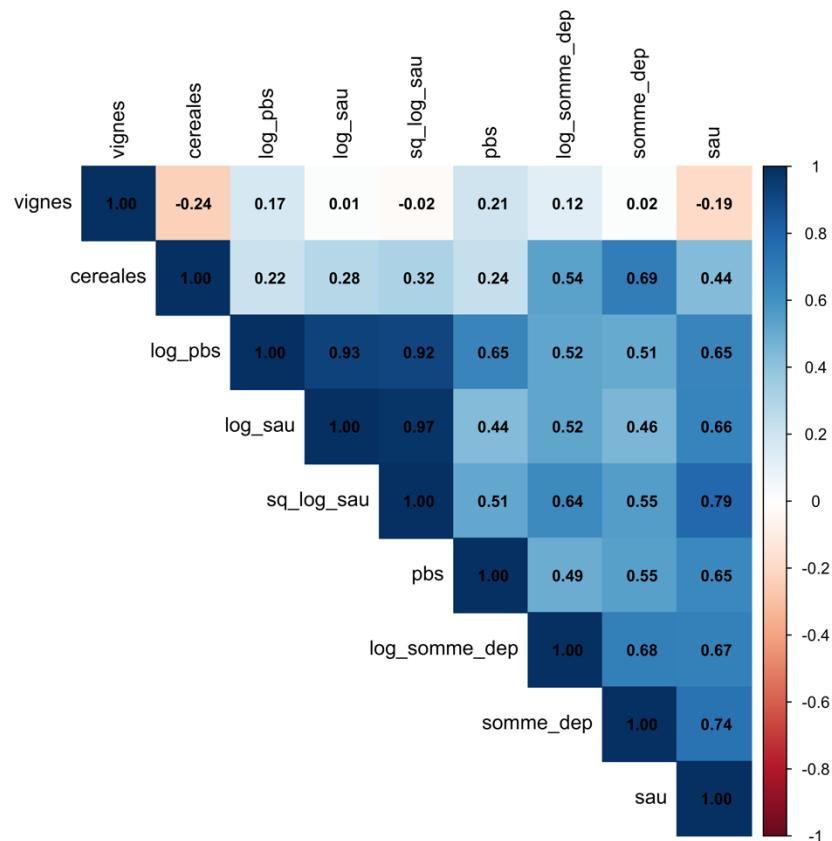
TOP 3 DES DÉPARTEMENTS LES PLUS VITICOLES		
Département	Part vignes (%)	SAU totale (ha)
Gironde	50.1	233 394
Hérault	44.7	175 790
Vaucluse	44.4	112 888

Source : Agreste - Recensement agricole 2020 | SAU en hectares

Dans ces tableaux, on voit le top 3 des départements qui achètent le plus de glyphosate ainsi que les départements qui ont le plus de céréales et les départements qui ont la plus grande production viticole. On voit donc que ce ne sont aucun des départements les plus céréaliers et les plus viticoles qui sont les départements qui achètent le plus de glyphosate. Bien que nous pensions jusque-là, au vu des résultats, que les déterminants de la quantité de glyphosate achetée étaient la production viticole et la production céréalière, la réalité est plus nuancée et on constate que même si ce sont des facteurs importants, d'autres facteurs déterminent aussi de manière importante la quantité de glyphosate achetée.

# Statistiques bivariées

## Correlations entre les variables

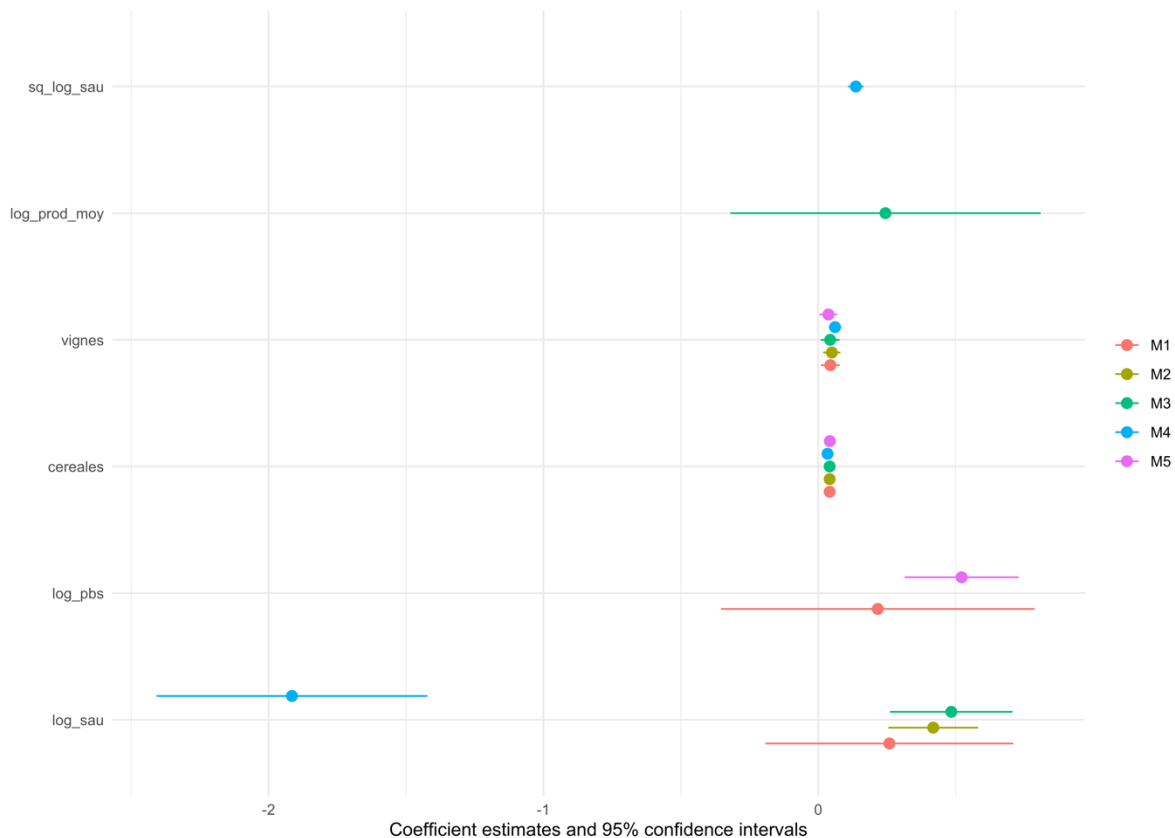


Nous remarquons aussi que la corrélation entre la `log_pbs` (logarithme de la Production Brute Standard) et la `log_sau` (logarithme de la Surface Agricole Utilisée) est très forte (0,92), ce qui est facile à envisager puisque la production brute standard dépend inévitablement de la surface cultivée (le seul cas contraire serait une terre avec un rendement tel qu'elle suffirait à substituer des hectares supplémentaires, ce qui serait une poule aux œufs d'or, mais entre nous, personne n'en a jamais vu...). C'est aussi le cas d'une corrélation forte (0,98) entre la variable `log_sau` et `sq_log_sau`, ce qui est évident puisque la variable `sq_log_sau` est juste le carré de la variable `log_sau`. Ensuite, nous retrouvons également une corrélation entre "somme\_dep", qui correspond à la quantité totale de glyphosate achetée dans le département, et la variable "cereales". Cette corrélation de 0,80 ne serait pas apparue si clairement si nous nous étions contentés des cartes vues plus haut. On peut donc écarter l'idée d'un overfitting.



## Test de plusieurs modèles

	M1	M2	M3	M4	M5
* p < 1, ** p < 0.05, *** p < 0.01					
(Intercept)	2.228*	3.124***	2.093*	10.454***	1.364*
	(1.561)	(1.008)	(1.568)	(1.046)	(1.366)
log_sau	0.259*	0.418***	0.484***	-1.915***	
	(0.227)	(0.082)	(0.112)	(0.248)	
log_pbs	0.216*				0.522***
	(0.287)				(0.104)
cereales	0.041***	0.041***	0.041***	0.034***	0.042***
	(0.007)	(0.007)	(0.007)	(0.005)	(0.007)
vignes	0.044**	0.049***	0.043**	0.061***	0.037**
	(0.017)	(0.016)	(0.017)	(0.011)	(0.016)
log_prod_moy			0.244*		
			(0.284)		
sq_log_sau				0.137***	
				(0.014)	
Num.Obs.	101	101	101	101	101
R2	0.496	0.493	0.497	0.743	0.489
R2 Adj.	0.475	0.477	0.476	0.733	0.473
AIC	380.4	379.0	380.3	312.3	379.8
BIC	396.1	392.1	395.9	328.0	392.9
Log.Lik.	-184.215	-184.512	-184.125	-150.146	-184.893
F	23.622	31.447	23.707	69.496	30.968
RMSE	1.50	1.50	1.50	1.07	1.51



D'après le  $R^2$  ajusté, les critères AIC et BIC, tous nous portent à croire que le modèle 4 est bien meilleur que les 4 autres.

## Notre meilleur modèle à estimer (MCO4)

$$\log(\text{somme\_dep}_i) = \text{const} + \beta_1 \log(\text{sau}_i) + \beta_2 \log(\text{sau}_i)^2 + \beta_3(\text{cereales}_i) + \beta_4(\text{vignes}_i) + u_i$$

On cherche à expliquer la quantité de glyphosate en fonction de la surface agricole exploitée, de la production brute standard, de la part de céréales dans l'exploitation, de la part de vignes dans l'exploitation. Le but étant de déterminer si la surface exploitée a un impact significatif sur la quantité de substance utilisée (ici le glyphosate), ainsi que déterminer si la culture, que ce soit les céréales ou les vignes, a un impact sur les quantités de glyphosate utilisés (savoir si en moyenne on utilise plus de glyphosate dans une culture que dans une autres).

## Meilleur modèle estimé et interprétation des paramètres

$$\log(\widehat{\text{somme\_dep}}_i) = 10.454 - 1.915 \log(\text{sau}_i) + 0.137 \log(\text{sau}_i)^2 + 0.034 \text{cereales}_i + 0.061 \text{vignes}_i$$

Lorsque tout les paramètres sont égaux à 0, le  $\log\_somme\_dep$  est de 10,454.

$$\frac{\partial \log(\widehat{\text{somme\_dep}}_i)}{\partial \log(\text{sau}_i)} = -1.915 + 2 \times 0.137 \log(\text{sau}_i) \Rightarrow -1.641 \log(\text{sau}_i)$$

Lorsque la SAU augmente de 1%, la quantité de glyphosate utilisée baisse de 1,641%. Puis, elle augmente à partir d'un certain seuil, d'où l'effet quadratique. Lorsque la part de céréales dans la SAU augmente de 1%, la quantité de glyphosate achetée augmente de 3.4%. Lorsque la part de vignes dans la SAU augmente de 1%, la quantité de glyphosate utilisée augmente de 6.1%.

Tous les paramètres de notre modèle sont significatifs au seuil de 99%.

$$\log(sau_i) = \frac{\beta_1}{2 \times \beta_2} = 7.00$$

$$\text{Donc } sau_i = \exp(6.989) = 1095$$

```
## L'effet marginal devient nul à partir de : log(sau) = 7 soit sau = 1095
hectares
```

La valeur de 1095 hectares nous semble peu réaliste, on a plutôt l'impression que le point où l'effet marginal devient nul est hors de notre jeu de données.

## Qualité d'ajustement

### Test de Fisher

$$H_0 : \hat{\beta}_k = 0 \quad k \in \{1, \dots, 4\}$$

$$H_1 : \hat{\beta}_k \neq 0 \quad k \in \{1, \dots, 4\}$$

Règle de décision : p-value < 0,05, rejet de H0, le modèle est globalement significatif.

```
## Test de Fisher pour la significativité globale du modèle :
## Statistique F = 69.5
## Degrés de liberté (modèle) = 4
## Degrés de liberté (résidus) = 96
## p-value = < 0.001
```

Test de Fisher - Significativité globale du modèle

	Statistique.F	ddl1	ddl2	p.value
value	69.5	4	96	< 0.001

P-value < 0,05, le modèle est globalement significatif.

## Analyse du R2

Le R2 du modèle MCO 4 (notre meilleur modèle) est de 0,733, autrement dit, 73,3% de la variance est expliquée par le modèle.

# Respect des hypothèses H1 à H5.

## H1 L'estimation de l'erreur est nulle

$$E[u] = 0$$

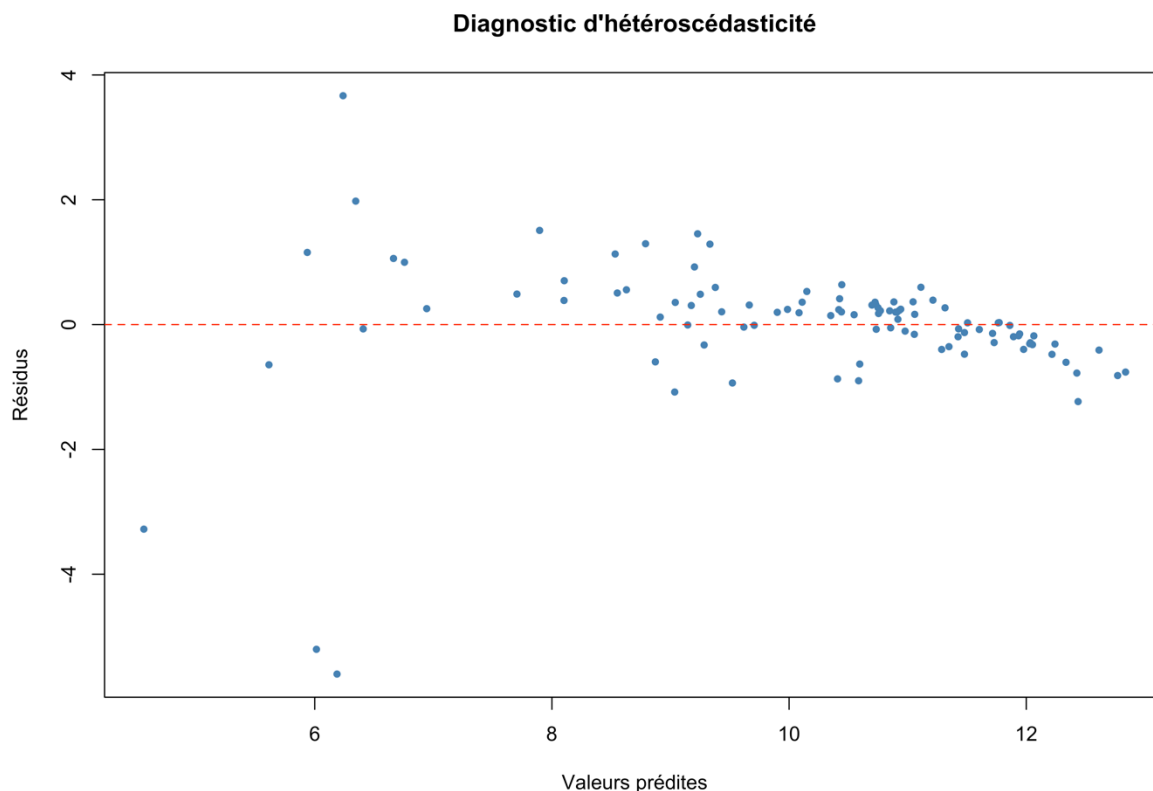
```
## One Sample t-test
##
## data:  res1
## t = 1.5683e-16, df = 100, p-value = 1
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -0.2122847  0.2122847
## sample estimates:
## mean of x
## 1.678044e-17
```

H1 est bien respecté, l'espérance du terme d'erreur n'est pas significativement différente de 0.

## H2 La variance de l'erreur est constante (Test d'hétéroscedasticité)

$$V[u] = \sigma^2$$

Visualisation de l'hétéroscedasticité



D'après ce graphique, on peut penser que l'on est en présence d'hétéroscedasticité.

## Test par la méthode de Breusch-Pagan

```
## Test de Breusch-Pagan :
##
## studentized Breusch-Pagan test
##
## data: modele_mco4
## BP = 24.225, df = 4, p-value = 7.2e-05
```

Conclusion On peut conclure de ce test qu'il y a présence d'hétéroscédasticité.

## Correction de l'hétéroscédasticité si constat d'hétéroscédasticité Comparaisons Écarts Types robuste et FGLS

Comparaison des trois méthodes d'estimation

	MCO Standard	MCO Robustes	FGLS
Constante	10.454***	10.454***	9.380***
	(1.046)	(1.564)	(0.788)
log(SAU)	-1.915***	-1.915***	-1.290***
	(0.248)	(0.451)	(0.183)
sq_log_sau	0.137***	0.137***	0.098***
	(0.014)	(0.026)	(0.011)
Part céréales (%)	0.034***	0.034***	0.027***
	(0.005)	(0.006)	(0.002)
Part vignes (%)	0.061***	0.061***	0.042***
	(0.011)	(0.013)	(0.007)
Num.Obs.	101	101	101
R2	0.743		0.802
R2 Adj.	0.733		0.794

+ p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

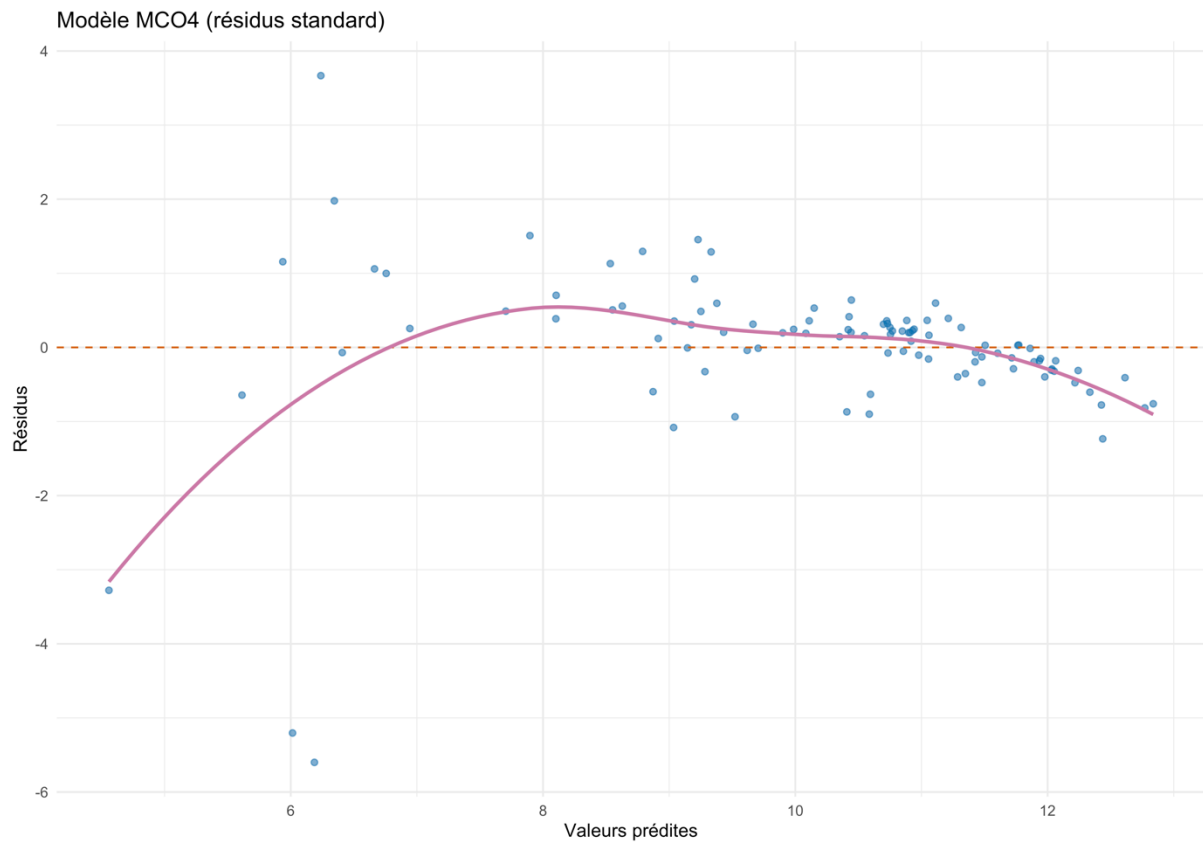
Erreurs standards robustes (HC3) pour la colonne 'MCO Robustes'

FGLS utilise une pondération par l'inverse de la variance estimée

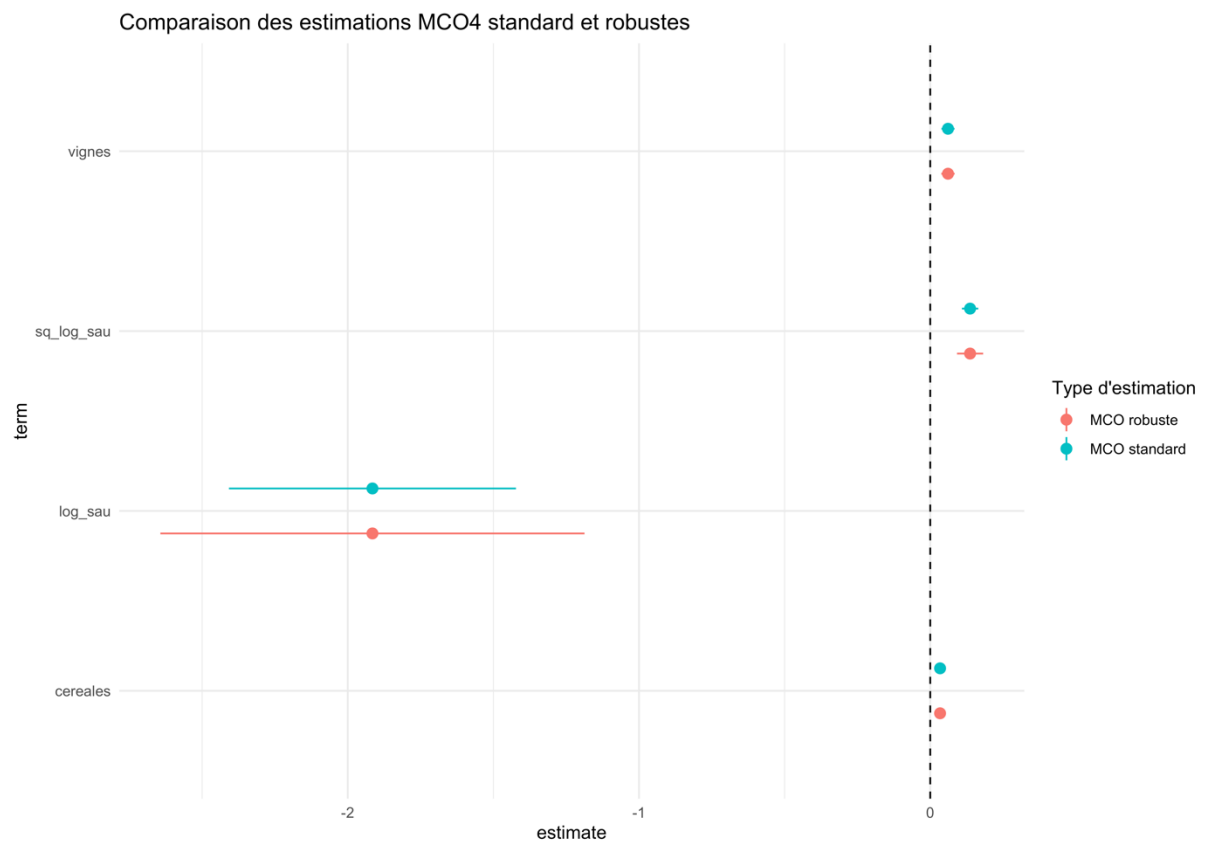
Les corrections par les écarts-types robustes modifient légèrement la significativité de nos paramètres mais ne changent pas les seuils dans notre modèle. Notre analyse reste sensiblement la même. La méthode des FGLS est meilleure que la méthode des MCO robuste car son  $R^2$  ajusté est plus fiable.

## Comparaison des modèles MCO4 vs FGLS

### Visualisation des modèles

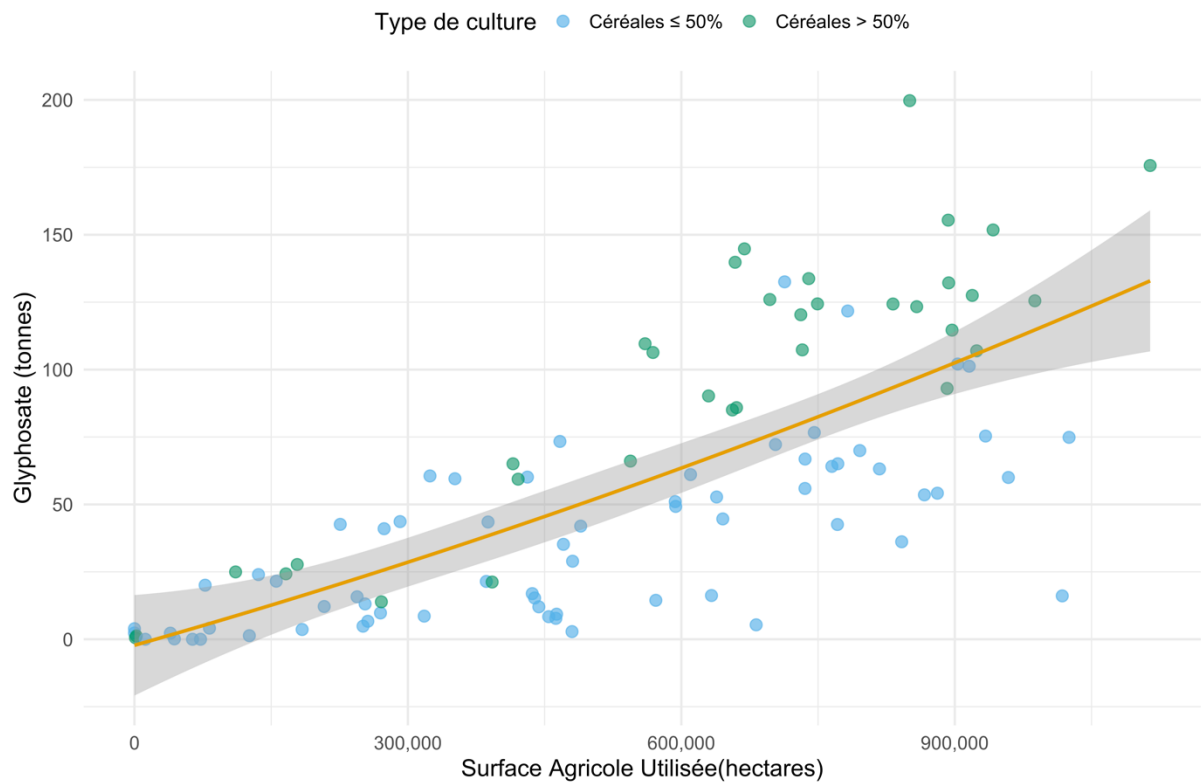


## Visualisation du modèle MCO4 robuste et comparaison avec MCO4 classique



La correction avec les FGLS nous montre aussi que la correction de l'hétéroscédasticité est bien présente, mais qu'il n'y a pas de valeur extrême. Donc, la méthode des FGLS est mieux pour notre modèle.

## Utilisation de glyphosate par Surface Agricole Utilisée

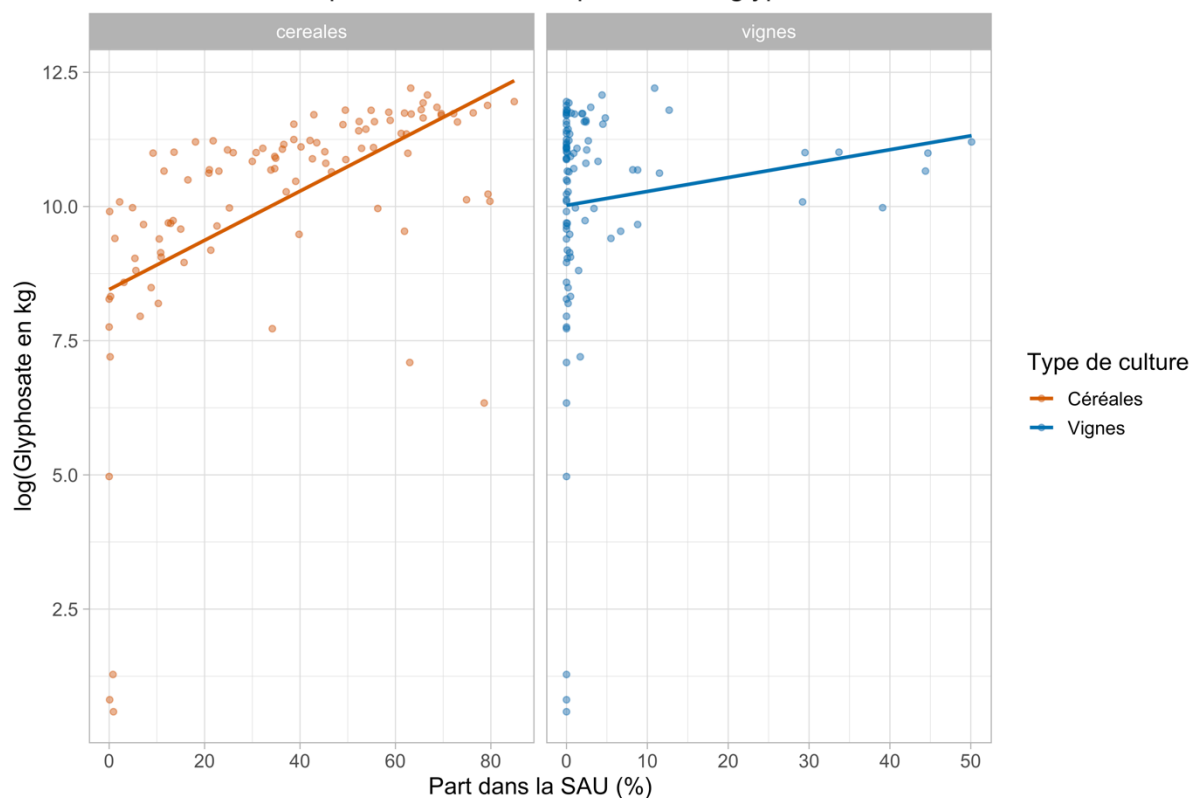


Ce graphique met en évidence la relation croissante entre la Surface agricole utilisée avec plus de 50% de céréale et la quantité de glyphosate achetée. On constate une certaine dispersion non linéaire entre ces valeurs, ce qui suggère d'autres facteurs explicatifs sur la quantité de glyphosate achetée. D'autre part, on constate aussi qu'à Surface égale, les départements avec une part de céréale plus faible achètent moins de glyphosate.

```
## `geom_smooth()` using formula = 'y ~ x'
```

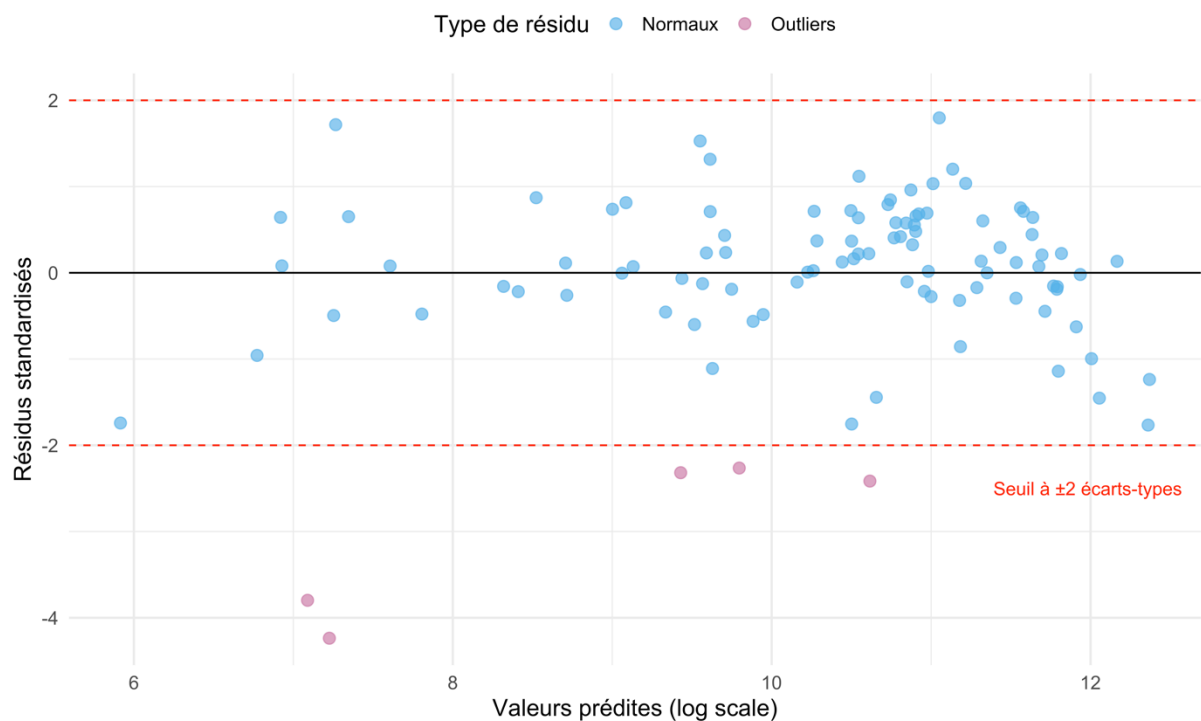


## Effet des cultures spécialisées sur les quantités de glyphosate utilisé



## Diagnostic des résidus du modèle FGLS

Après correction de l'hétéroscédasticité



Test de Breusch-Pagan  $p = 0$

H3 La matrice X est non aléatoire, la variable est non aléatoire pour toutes les observations

Dans notre modèle, on pourrait être sujet à un problème de variables endogènes.

- La variable SAU pourrait être corrélée avec notre terme d'erreur. Des événements extérieurs pourraient venir affecter en même temps l'utilisation de glyphosate et la SAU. On peut prendre comme exemple les conditions environnementales, la demande des utilisateurs, les caractéristiques économiques du marché ou encore les politiques agricoles.
- Les variables vignes et céréales pourraient être corrélées avec notre terme d'erreur. Dans un cas météorologique ou géographique où l'utilisation de glyphosate est plus accrue, la culture de vignes ou de céréales est meilleure. On peut aussi avoir des raisons d'agrandissement de culture de vignes qui entraîne une utilisation plus forte de glyphosate comme herbicide.

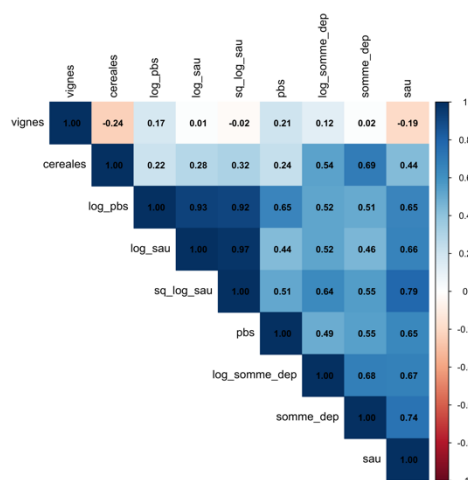
Dans le cas de notre modèle MCO4, la variable qui pourrait être sujette à l'endogénéité est la variable SAU. En effet, celle-ci est la plus sujette à ce problème, car elle ne suit pas nos intuitions de départ en étant négative, ce qui pourrait être expliqué par le fait qu'elle soit corrélée avec des facteurs non observés. Le problème est que notre base de données ne nous permet pas de faire de test pour ce problème d'endogénéité, car nous n'avons pas de variable instrumentale qui soit corrélée avec la SAU et non corrélée avec notre terme d'erreur.

#### H4 Le modèle est correctement spécifié

##	log_sau	sq_log_sau	cereales	vignes
##	15.668758	16.073588	1.127131	1.075180

Les résultats du VIF (Variance Inflation Factor) nous montrent que pour les variables céréales et vignes, il n'y a pas de problème de multicolinéarité, car leurs valeurs sont inférieures à 5. Les paramètres `log_sau` et `sq_log_sau`, en revanche, montrent un risque de multicolinéarité forte, car leur VIF est bien supérieur à 5, voire même supérieur à 10, ce qui indique une forte colinéarité. On en conclut donc que notre modèle n'est pas correctement spécifié, car nos variables `log_sau` et `sq_log_sau` sont multicolinéaires.

H5 La matrice  $X$  est de plein rang



La matrice est bien de plein rang, la colinéarité forte entre  $\log(sau_i)$  et  $\log(sau_i^2)$  est due au fait que c'est la même variable que nous avons transformée pour la mettre au carré afin de capter les relations quadratiques de cette variable explicative avec la variable dépendante.

# Modèle final

## Interprétation

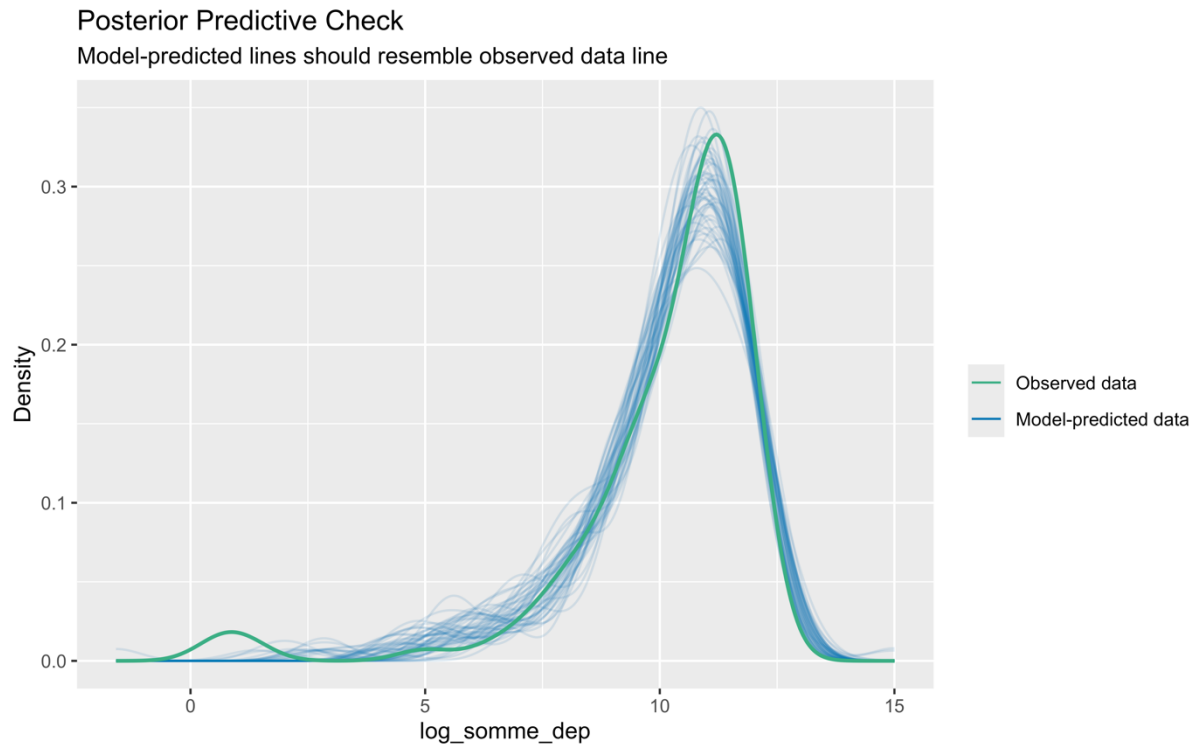
```
##
## =====
##                               Dependent variable:
##                               -----
##                               log_somme_dep
##                               -----
## log_sau                      -1.290***
##                               (0.183)
##
## sq_log_sau                   0.098***
##                               (0.011)
##
## cereales                    0.027***
##                               (0.002)
##
## vignes                      0.042***
##                               (0.007)
##
## Constant                     9.380***
##                               (0.788)
##
## -----
## Observations                 101
## R2                          0.802
## Adjusted R2                 0.794
## Residual Std. Error        1.719 (df = 96)
## F Statistic                 97.451*** (df = 4; 96)
## =====
## Note:                       *p<0.1; **p<0.05; ***p<0.01
```

$$\widehat{\log(\text{somme\_dep}_i)} = 9.380 - 1.29 \log(\text{sau}_i) + 0.098 \log(\text{sau}_i)^2 + 0.027 \text{ cereales}_i + 0.042 \text{ vignes}_i$$

$$\frac{\partial \widehat{\log(\text{somme\_dep}_i)}}{\partial \log(\text{sau}_i)} = -1.915 + 2 \times 0.137 \log(\text{sau}_i) \Rightarrow -1.641 \log(\text{sau}_i)$$

Lorsque tout les paramètres sont égaux à 0, le log\_somme\_dep est de 9.38. Lorsque la SAU augmente de 1%, la quantité de glyphosate utilisée baisse de 1,094%. Puis, elle augmente à partir d'un certain seuil, d'où l'effet quadratique. Lorsque la part de céréales dans la SAU augmente de 1%, la quantité de glyphosate achetée augmente de 2.7%. Lorsque la part de vignes dans la SAU augmente de 1%, la quantité de glyphosate utilisée augmente de 4.2%. La part de la variance expliquée par le modèle est de 79.4%. Tous les paramètres de notre modèle sont significatifs au seuil de 99%.

## Performance du modèle



Ce graphique nous montre que les estimations que nous avons faites se rapprochent des valeurs observées. On en conclut que notre modèle est plutôt performant.

## Conclusion

On conclut que toutes les variables sont significatives. Les facteurs observés dans notre analyse sont tous responsables de la détermination de la quantité de glyphosate, cependant leur importance est inégale. Après recherche et implémentation de différents modèles, notre modèle 4 est le plus pertinent et le plus correct selon les critères AIC et BIC. On voit que toutes les variables sont significatives, c'est-à-dire que tous les paramètres ont leur importance dans la détermination de la quantité de glyphosate utilisée. Cependant, il semble, comme le montrent ces résultats et le test de l'hypothèse H4, qu'il manque des variables importantes, bien que le  $R^2$  soit de 79,4%. Cependant, nous n'avons pas de multicolinéarité. Le modèle est performant et nous permet d'affirmer avec une certaine certitude que les cultures céréales et vignes, ainsi que la surface agricole utilisée, ont tous un impact dans la détermination de la quantité de glyphosate achetée.